

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering the required information, reviewing and collecting the information, and completing the review of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Project Director, Washington, DC 20503.

Reviewing
Information

0697

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE May, 1996		3. REPORT NUMBER	
4. TITLE AND SUBTITLE 1995 Summer Research Program (SRP), Summer Research Extension Program (SREP), Final Report, Volume 2, Phillips Laboratory				5. FUNDING NUMBERS F49620-93-C-0063	
6. AUTHOR(S) Gary Moore					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research & Development Laboratories (RDL) 5800 Uplander Way Culver City, CA 90230-6608				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research (AFOSR) 801 N. Randolph St. Arlington, VA 22203-1977				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The United States Air Force Summer Research Program (SRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members, graduate students, and high school students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates and Air Force Air Logistics Centers (ALC). AFOSR also offers its research associates (faculty only) an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. This volume consists of the SREP program background, management information, statistics, a listing of the participants, and the technical report for each participant of the SREP working at the AF Phillips Laboratory.					
14. SUBJECT TERMS Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student				15. NUMBER OF PAGES	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL		

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1995
SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 2

PHILLIPS LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way
Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

Program Manager, RDL
Scott Licoscas

Program Administrator, RDL
Gwendolyn Smith

Program Administrator
Johnetta Thompson

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base
Washington, D.C.

May 1996

AQM01-06-0909

20010322 002

PREFACE

This volume is part of a five-volume set that summarizes the research of participants in the 1995 AFOSR Summer Research Extension Program (SREP). The current volume, Volume 1 of 5, presents the final reports of SREP participants at Armstrong Laboratory, Phillips Laboratory, Rome Laboratory, Wright Laboratory, Arnold Engineering Development Center, Frank J. Seiler Research Laboratory, and Wilford Hall Medical Center.

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3, with each series of reports preceded by a management summary. Reports in the five-volume set are organized as follows:

VOLUME	TITLE
1A	Armstrong Laboratory (part one)
1B	Armstrong Laboratory (part two)
2	Phillips Laboratory
3	Rome Laboratory
4A	Wright Laboratory (part one)
4B	Wright Laboratory (part two)
5	Arnold Engineering Development Center Frank J. Seiler Research Laboratory Wilford Hall Medical Center

1995 SREP FINAL REPORTS

Armstrong Laboratory

VOLUME 1

Report #	Report Title Author's University	Report Author
1	Determination of the Redox Capacity of Soil Sediment and Prediction of Pollutant University of Georgia, Athens, GA	Dr. James Anderson Analytical Chemistry AL/EQ
2	Finite Element Modeling of the Human Neck and Its Validation for the ATB Villanova University, Villanova, PA	Dr. Hashem Ashrafiuon Mechanical Engineering AL/CF
3	An Examination of the Validity of the Experimental Air Force ASVAB Composites Tulane University, New Orleans, LA	Dr. Michael Burke Psychology AL/HR
4	Fuel Identification by Neural Networks Analysis of the Response of Vapor Sensitive Sensors Arrays Edinboro University of Pennsylvania, Edinboro, PA	Dr. Paul Edwards Chemistry AL/EQ
5	A Comparison of Multistep vs Singlestep Arrhenius Integral Models for Describing Laser Induced Thermal Damage Florida International University, Miami, FL	Dr. Bernard Gerstman Physics AL/OE
6	Effects of Mental Workload and Electronic Support on Negotiation Performance University of Dayton, Dayton, OH	Dr. Kenneth Graetz Psychology AL/HR
7	Regression to the Mean in Half Life Studies University of Main, Orono, ME	Dr. Pushpa Gupta Mathematics & Statistics AL/AO
8	Application of the MT3D Solute Transport Model to the Made-2 Site: Calibration Florida State University, Tallahassee, FL	Dr. Manfred Koch Geophysics AL/EQ
9	Computer Calculations of Gas-Phase Reaction Rate Constants Florida State University, Tallahassee, FL	Dr. Mark Novotny SupercompComp. Res. I AL/EQ
10	Surface Fitting Three Dimensional Human Scan Data Ohio University, Athens, OH	Dr. Joseph Nurre Mechanical Engineering AL/CF
11	The Effects of Hyperbaric Oxygenation on Metabolism of Drugs and Other Xenobioti University of So. Carolina, Columbia, So. Carolina	Dr. Edward Piepmeier Pharmaceutics AL/AO
12	Maintaining Skills After Training: The Role of Opportunity to Perform Trained Tasks on Training Effectiveness Rice University, Houston, TX	Dr. Miguel Quinones Psychology AL/HR

1995 SREP FINAL REPORTS

Armstrong Laboratory

VOLUME 1 (cont.)

Report #	Report Title Author's University	Report Author
13	Nonlinear Transcutaneous Electrical Stimulation of the Vestibular System University of Illinois Urbana-Champaign, Urbana,IL	Dr. Gary Riccio Psychology AL/CF
14	Documentation of Separating and Separated Boundary Layer Flow, For Application Texas A&M University, College Station, TX	Dr. Wayne Shebilske Psychology AL/HR
15	Tactile Feedback for Simulation of Object Shape and Textural Information in Haptic Displays Ohio State University, Columbus, OH	Dr. Janet Weisenberger Speech & Hearing AL/CF
16	Melatonin Induced Prophylactic Sleep as a Countermeasure for Sleep Deprivation Oregon Health Sciences University, Portland, OR	Mr. Rod Hughes Psychology AL/CF

1995 SREP FINAL REPORTS

Phillips Laboratory

VOLUME 2A

Report #	Report Title Author's University	Report Author
1	Investigation of the Mixed-Mode Fracture Behavior of Solid Propellants University of Houston, Houston, TX	Dr. K. Ravi-Chandar Aeronautics PL/RK
2	Performance Study of ATM-Satellite Network SUNY-Buffalo, Buffalo, NY	Dr. Nasser Ashgriz Mechanical Engineering PL/RK
3	Characterization of CMOS Circuits Using a Highly Calibrated Low-Energy X-Ray Source Embry-Riddle Aeronautical Univ., Prescott, AZ	Dr. Raymond Bellem Computer Science PL/VT
4	Neutron Diagnostics for Pulsed Plasmas of Compact Toroid-Marauder Type Stevens Institute of Tech, Hoboken, NJ	Dr. Jan Brzosko Nuclear Physics PL/WS
5	Parallel Computation of Zernike Aberration Coefficients for Optical Aberration Correction University of Houston-Victoria, Victoria, TX	Dr. Meledath Damodaran Math & Computer Science PL/LI
6	Quality Factor Evaluation of Complex Cavities University of Denver, Denver, CO	Dr. Ronald DeLyser Electrical Engineering PL/WS
7	Unidirectional Ring Lasers and Laser Gyros with Multiple Quantum Well Gain University of New Mexico, Albuquerque, NM	Dr. Jean-Claude Diels Physics PL/LI
8	A Tool for the Formation of Variable Parameter Inverse Synthetic Aperture Radar University of Nevada, Reno, NV	Dr. James Henson Electrical Engineering PL/WS
9	Radar Ambiguity Functionals Univ. of Massachusetts at Lowell, Lowell, MA	Dr. Gerald Kaiser Physics PL/GP
10	The Synthesis and Chemistry of Peroxonitrites Peroxonitrous Acid Univ. of Massachusetts at Lowell, Lowell, MA	Dr. Albert Kowalak Chemistry PL/GP
11	Temperature and Pressure Dependence of the Band Gaps and Band Offsets University of Houston, Houston, TX	Dr. Kevin Malloy Electrical Engineering PL/VT
12	Theoretical Studies of the Performance of Novel Fiber-Coupled Imaging Interferom University of New Mexico, Albuquerque, NM	Dr. Sudhakar Prasad Physics PL/LI

1995 SREP FINAL REPORTS

Phillips Laboratory

VOLUME 2B

Report #	Author's University	Report Author
13	Static and Dynamic Graph Embedding for Parallel Programming Texas AandM Univ.-Kingsville, Kingsville, TX	Dr. Mark Purtill Mathematics PL/WS
14	Ultrafast Process and Modulation in Iodine Lasers University of New Mexico, Albuquerque, NM	Dr. W. Rudolph Physics PL/LI
15	Impedance Matching and Reflection Minimization for Transient EM Pulses Through University of New Mexico, Albuquerque, NM	Dr. Alexander Stone Mathematics and Statics PL/WS
16	Low Power Retromodular Based Optical Transceiver for Satellite Communications Utah State University, Logan, UT	Dr. Charles Swenson Electrical Engineering PL/VT
17	Improved Methods of Tilt Measurement for Extended Images in the Presence of Atmospheric Disturbances Using Optical Flow Michigan Technological Univ., Houghton, MI	Mr. John Lipp Electrical Engineering PL/LI
18	Thermoluminescence of Simple Species in Molecular Hydrogen Matrices Cal State Univ.-Northridge, Northridge, CA	Ms. Janet Petroski Chemistry PL/RK
19	Design, Fabrication, Intelligent Cure, Testing, and Flight Qualification University of Cincinnati, Cincinnati, OH	Mr. Richard Salasovich Mechanical Engineering PL/VT

1995 SREP FINAL REPORTS

Rome Laboratory

VOLUME 3

Report #	Author's University	Report Author
1	Performance Study of an ATM/Satellite Network Florida Atlantic University, Boca Raton, FL	Dr. Valentine Aalo Electrical Engineering RL/C3
2	Interference Excision in Spread Spectrum Communication Systems Using Time-Frequency Distributions Villanova University, Villanova, PA	Dr. Moeness Amin Electrical Engineering RL/C3
3	Designing Software by Reformulation Using KIDS Oklahoma State University, Stillwater, OK	Dr. David Benjamin Computer Science RL/C3
4	Detection Performance of Over Resolved Targets with Non-Uniform and Non-Gaussian Howard University, Washington, DC	Dr. Ajit Choudhury Engineering RL/OC
5	Computer-Aided-Design Program for Solderless Coupling Between Microstrip and Stripline Structures Southern Illinois University, Carbondale, IL	Dr. Frances Harackiewicz Electrical Engineering RL/ER
6	Spanish Dialect Identification Project Colorado State University, Fort Collins, CO	Dr. Beth Losiewicz Psycholinguistics RL/IR
7	Automatic Image Registration Using Digital Terrain Elevation Data University of Maine, Orono, ME	Dr. Mohamed Musavi Engineering RL/IR
8	Infrared Images of Electromagnetic Fields University of Colorado, Colorado Springs, CO	Dr. John Norgard Engineering RL/ER
9	Femtosecond Pump-Probe Spectroscopy System SUNY Institute of Technology, Utica, NY	Dr. Dean Richardson Photonics RL/OC
10	Synthesis and Properties B-Diketonate-Modified Heterobimetallic Alkoxides Tufts University, Medford, MA	Dr. Daniel Ryder, Jr. Chemical Engineering RL/ER
11	Optoelectronic Study of Semiconductor Surfaces and Interfaces Rensselaer Polytechnic Institute, Troy, NY	Dr. Xi-Cheng Zhang Physics RL/ER

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4A

Report #	Author's University	Report Author
1	An Investigation of the Heating and Temperature Distribution in Electrically Excited Foils Auburn University, Auburn, AL	Dr. Michael Baginski Electrical Engineering WL/MN
2	Micromechanics of Creep in Metals and Ceramics at High Temperature Wayne State University, Detroit, MI	Dr. Victor Berdichevsky Aerospace Engineering WL/FI
3	Development of a Fluorescence-Based Chemical Sensor for Simultaneous Oxygen Quantitation and Temp. Measurement Columbus College, Columbus, GA	Dr. Steven Buckner Chemistry WL/PO
4	Development of High-Performance Active Dynamometer Sys. for Machines and Drive Clarkson University, Potsdam, NY	Dr. James Carroll Electrical Engineering WL/PO
5	SOLVING $z(t)=1n[Acos(w_1t)+Bcos(w_2)+C]$ Transylvania University, Lexington, KY	Dr. David Choate Mathematics WL/AA
6	Synthesis, Processing and Characterization of Nonlinear Optical Polymer Thin Films University of Cincinnati, Cincinnati, OH	Dr. Stephen Clarson Mats Science & Engineering WL/ML
7	An Investigation of Planning and Scheduling Algorithms for Sensor Management Embry-Riddle Aeronautical University, Prescott, AZ	Dr. Milton Cone Comp. Science & Engineering WL/AA
8	A Study to Determine Wave Gun Firing Cycles for High Performance Model Launches Louisiana State University, Baton Rouge, LA	Dr. Robert Courter Mechanical Engineering WL/MN
9	Characterization of Electro-Optic Polymers University of Dayton, Dayton, OH	Dr. Vincent Dominic Electro Optics Program WL/ML
10	A Methodology for Affordability in the Design Process Clemson University, Clemson, SC	Dr. Georges Fadel Mechanical Engineering WL/MT
11	Data Reduction and Analysis for Laser Doppler Velocimetry North Carolina State University, Raleigh, NC	Dr. Richard Gould Mechanical Engineering WL/PO

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4A (cont.)

Report #	Author's University	Report Author
12	Hyperspectral Target Identification Using Bomen Spectrometer Data University of Dayton, Dayton, OH	Dr. Russell Hardie Electrical Engineering WL/AA
13	Robust Fault Detection and Classification Auburn University, Auburn, AL	Dr. Alan Hodel Electrical Engineering WL/MN
14	Multidimensional Algorithm Development and Analysis Mississippi State University, Mississippi State University, MS	Dr. Jonathan Janus Aerospace Engineering WL/MN
15	Characterization of Interfaces in Metal-Matrix Composites Michigan State University, East Lansing, MI	Dr. Iwona Jasiuk Materials Science WL/ML
16	TSI Mitigation: A Mountaintop Database Study Lafayette College, Easton, PA	Dr. Ismail Jouny Electrical Engineering WL/AA
17	Comparative Study and Performance Analysis of High Resolution SAR Imaging Techniques University of Florida, Gainesville, FL	Dr. Jian Li Electrical Engineering WL/AA

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4B

Report #	Author's University	Report Author
18	Prediction of Missile Trajectory University of Missouri-Columbia, Columbia, MO	Dr. Chun-Shin Lin Electrical Engineering WL/FI
19	Three Dimensional Deformation Comparison Between Bias and Radial Aircraft Tires Cleveland State University, Cleveland, OH	Dr. Paul Lin Mechanical Engineering WL/FI
20	Investigation of AlGaAs/GaAs Heterojunctin Bipolar Transistor Reliability Based University of Central Florida, Orlando, FL	Dr. Juin Liou Electrical Engineering WL/EL
21	Thermophysical Invariants From LWIR Imagery for ATR University of Virginia, Charlottesville, VA	Dr. Nagaraj Nandhakumar Electrical Engineering WL/AA
22	Effect of Electromagnetic Environment on Array Signal Processing University of Dayton, Dayton, OH	Dr. Krishna Pasala Electrical Engineering WL/AA
23	Functional Decomposition of Binary, Multiple-Valued, and Fuzzy Logic Portland State University, Portland, OR	Dr. Marek Perkowski Electrical Engineering WL/AA
24	Superresolution of Passive Millimeter-Wave Imaging Auburn University, Auburn, AL	Dr. Stanley Reeves Electrical Engineering WL/MN
25	Development of a Penetrator Optimizer University of Alabama, Tuscaloosa, AL	Dr. William Rule Engineering Science WL/MN
26	Heat Transfer for Turbine Blade Film Cooling with Free Stream Turbulence-Measurements and Predictions University of Dayton, Dayton, OH	Dr. John Schauer Mech. & Aerosp. Engineering WL/FI
27	Neural Network Identification and Control in Metal Forging University of Florida, Gainesville, FL	Dr. Carla Schwartz Electrical Engineering WL/FI
28	Documentation of Separating and Separated Boundary Layer Flow, for Application University of Minnesota, Minneapolis, MN	Dr. Terrence Simon Mechanical Engineering WL/PO
29	Transmission Electron Microscopy of Semiconductor Heterojunctions Carnegie Melon University, Pittsburgh, PA	Dr. Marek Skowronski Matls Science & Engineering WL/EL

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4B (cont.)

Report #	Author's University	Report Author
30	Parser in SWI-PROLOG Wright State University, Dayton, OH	Dr. K. Thirunarayan Computer Science WL/EL
31	Development of Qualitative Process Control Discovery Systems for Polymer Composite and Biological Materials University of California, Los Angeles, CA	Dr. Robert Trelease Anatomy & Cell Biology WL/ML
32	Improved Algorithm Development of Massively Parallel Epic Hydrocode in Cray T3D Massively Parallel Computer Florida Atlantic University, Boca Raton, FL	Dr. Chi-Tay Tsai Engineering Mechanics WL/MN
33	The Characterization of the Mechanical Properties of Materials in a Biaxial Stress Environment University of Kentucky, Lexington, KY	Dr. John Lewis Materials Science Engineering WL/MN

1995 SREP FINAL REPORTS

VOLUME 5

Report #	Author's University	Report Author
Arnold Engineering Development Center		
1	Plant-Wide Preventive Maintenance and Monitoring Vanderbilt University	Mr. Theodore Bapty Electrical Engineering AEDC
Frank J. Seiler Research Laboratory		
1	Block Copolymers at Inorganic Solid Surfaces Colorado School of Mines, Golden, CO	Dr. John Dorgan Chemical Engineering FJSRL
2	Non-Linear Optical Properties of Polyacetylenes and Related Barry University, Miami, FL	Dr. M. A. Jungbauer Chemistry FJSRL
3	Studies of Second Harmonic Generation in Glass Waveguides Allegheny College, Meadville, PA	Dr. David Statman Physics FJSRL
Wilford Hall Medical Center		
1	Biochemical & Cell Physiological Aspects of Hyperthermia University of Miami, Coral Gables, FL	Dr. W. Drost-Hansen Chemistry WHMC

1995 SUMMER RESEARCH EXTENSION PROGRAM (SREP) MANAGEMENT REPORT

1.0 BACKGROUND

Under the provisions of Air Force Office of Scientific Research (AFOSR) contract F49620-90-C-0076, September 1990, Research & Development Laboratories (RDL), an 8(a) contractor in Culver City, CA, manages AFOSR's Summer Research Program. This report is issued in partial fulfillment of that contract (CLIN 0003AC).

The Summer Research Extension Program (SREP) is one of four programs AFOSR manages under the Summer Research Program. The Summer Faculty Research Program (SFRP) and the Graduate Student Research Program (GSRP) place college-level research associates in Air Force research laboratories around the United States for 8 to 12 weeks of research with Air Force scientists. The High School Apprenticeship Program (HSAP) is the fourth element of the Summer Research Program, allowing promising mathematics and science students to spend two months of their summer vacations working at Air Force laboratories within commuting distance from their homes.

SFRP associates and exceptional GSRP associates are encouraged, at the end of their summer tours, to write proposals to extend their summer research during the following calendar year at their home institutions. AFOSR provides funds adequate to pay for SREP subcontracts. In addition, AFOSR has traditionally provided further funding, when available, to pay for additional SREP proposals, including those submitted by associates from Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs). Finally, laboratories may transfer internal funds to AFOSR to fund additional SREPs. Ultimately the laboratories inform RDL of their SREP choices, RDL gets AFOSR approval, and RDL forwards a subcontract to the institution where the SREP associate is employed. The subcontract (see Appendix 1 for a sample) cites the SREP associate as the principal investigator and requires submission of a report at the end of the subcontract period.

Institutions are encouraged to share costs of the SREP research, and many do so. The most common cost-sharing arrangement is reduction in the overhead, fringes, or administrative charges institutions would normally add on to the principal investigator's or research associate's labor. Some institutions also provide other support (e.g., computer run time, administrative assistance, facilities and equipment or research assistants) at reduced or no cost.

When RDL receives the signed subcontract, we fund the effort initially by providing 90% of the subcontract amount to the institution (normally \$18,000 for a \$20,000 SREP). When we receive the end-of-research report, we evaluate it administratively and send a copy to the laboratory for a technical evaluation. When the laboratory notifies us the SREP report is acceptable, we release the remaining funds to the institution.

2.0 THE 1995 SREP PROGRAM

SELECTION DATA: A total of 719 faculty members (SFRP Associates) and 286 graduate students (GSRP associates) applied to participate in the 1994 Summer Research Program. From these applicants 185 SFRPs and 121 GSRPs were selected. The education level of those selected was as follows:

1994 SRP Associates, by Degree			
SFRP		GSRP	
PHD	MS	MS	BS
179	6	52	69

Of the participants in the 1994 Summer Research Program 90 percent of SFRPs and 25 percent of GSRPs submitted proposals for the SREP. Ninety proposals from SFRPs and ten from GSRPs were selected for funding, which equates to a selection rate of 54% of the SFRP proposals and of 34% for GSRP proposals.

1995 SREP: Proposals Submitted vs. Proposals Selected			
	Summer 1994 Participants	Submitted SREP Proposals	SREPs Funded
SFRP	185	167	90
GSRP	121	29	10
TOTAL	306	196	100

The funding was provided as follows:

Contractual slots funded by AFOSR	75
Laboratory funded	14
Additional funding from AFOSR	<u>11</u>
Total	100

Six HBCU/MI associates from the 1994 summer program submitted SREP proposals; six were selected (none were lab-funded; all were funded by additional AFOSR funds).

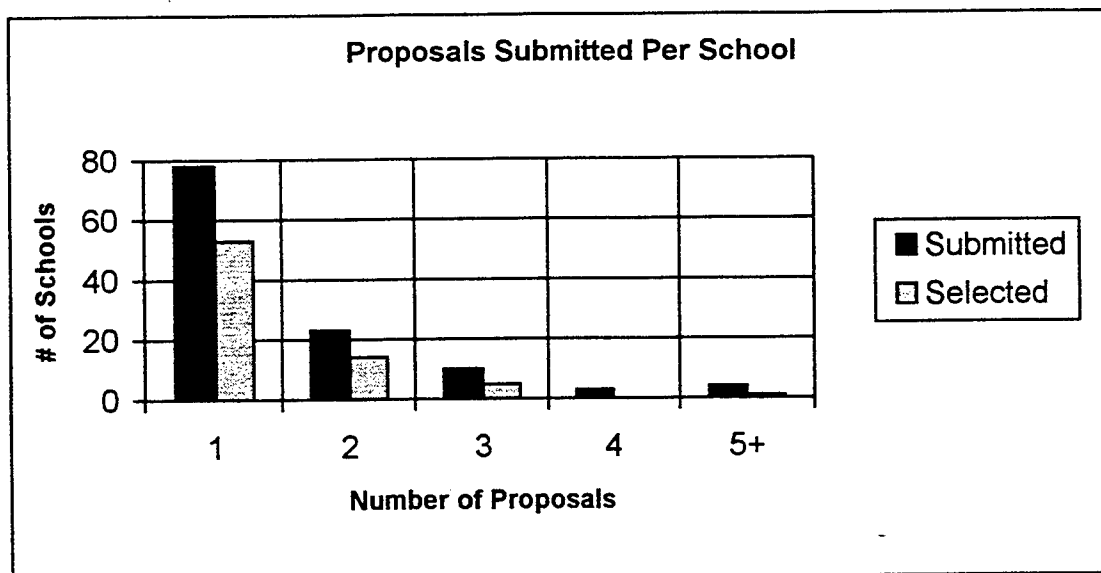
Proposals Submitted and Selected, by Laboratory		
	Applied	Selected
Armstrong Laboratory	41	19
Arnold Engineering Development Center	12	4
Frank J. Seiler Research Laboratory	6	3
Phillips Laboratory	33	19
Rome Laboratory	31	13
Wilford Hall Medical Center	2	1
Wright Laboratory	62	37
TOTAL		

Note: Phillips Laboratory funded 3 SREPs; Wright Laboratory funded 11; and AFOSR funded 11 beyond its contractual 75.

The 306 1994 Summer Research Program participants represented 135 institutions.

Institutions Represented on the 1994 SRP and 1995 SREP		
Number of schools represented in the Summer 92 Program	Number of schools represented in submitted proposals	Number of schools represented in Funded Proposals
135	118	73

Forty schools had more than one participant submitting proposals.



The selection rate for the 78 schools submitting 1 proposal (68%) was better than those submitting 2 proposals (61%), 3 proposals (50%), 4 proposals (0%) or 5+ proposals (25%). The 4 schools that submitted 5+ proposals accounted for 30 (15%) of the 196 proposals submitted.

Of the 196 proposals submitted, 159 offered institution cost sharing. Of the funded proposals which offered cost sharing, the minimum cost share was \$1000.00, the maximum was \$68,000.00 with an average cost share of \$12,016.00.

Proposals and Institution Cost Sharing		
	Proposals Submitted	Proposals Funded
With cost sharing	159	82
Without cost sharing	37	18
Total	196	100

The SREP participants were residents of 41 different states. Number of states represented at each laboratory were:

States Represented, by Proposals Submitted/Selected per Laboratory		
	Proposals Submitted	Proposals Funded
Armstrong Laboratory	21	13
Arnold Engineering Development Center	5	2
Frank J. Seiler Research Laboratory	5	3
Phillips Laboratory	16	14
Rome Laboratory	14	7
Wilford Hall Medical Center	2	1
Wright Laboratory	24	20

Eleven of the 1995 SREP Principal Investigators also participated in the 1994 SREP.

ADMINISTRATIVE EVALUATION: The administrative quality of the SREP associates' final reports was satisfactory. Most complied with the formatting and other instructions provided to them by RDL. Ninety seven final reports and two interim reports have been received and are included in this report. The subcontracts were funded by \$1,991,623.00 of Air Force money. Institution cost sharing totaled \$985,353.00.

TECHNICAL EVALUATION: The form used for the technical evaluation is provided as Appendix 2. ninety-two evaluation reports were received. Participants by laboratory versus evaluations submitted is shown below:

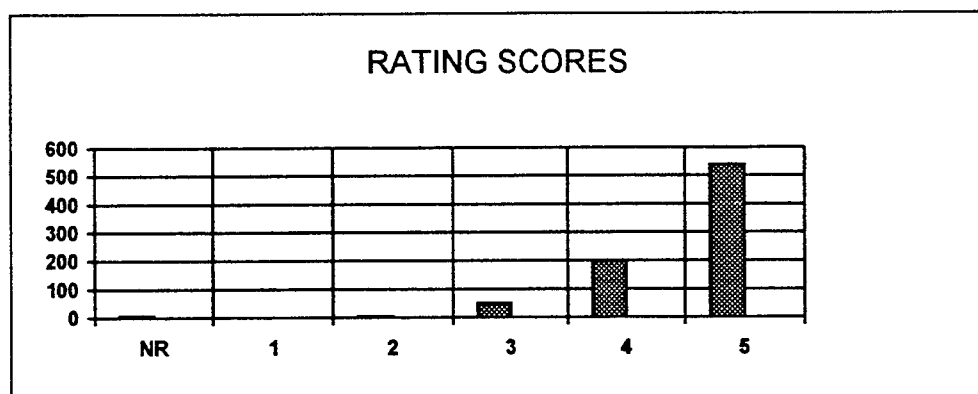
	Participants	Evaluations	Percent
Armstrong Laboratory	23 ¹	20	95.2
Arnold Engineering Development Center	4	4	100
Frank J. Seiler Research Laboratory	3	3	100
Phillips Laboratory	19 ²	18	100
Rome Laboratory	13	13	100
Wilford Hall Medical Center	1	1	100
Wright Laboratory	37	34	91.9
Total			

Notes:

- 1: Research on two of the final reports was incomplete as of press time so there aren't any technical evaluations on them to process, yet. Percent complete is based upon 20/21=95.2%
- 2: One technical evaluation was not completed because one of the final reports was incomplete as of press time. Percent complete is based upon 18/18=100%
- 3: See notes 1 and 2 above. Percent complete is based upon 93/97=95.9%

The number of evaluations submitted for the 1995 SREP (95.9%) shows a marked improvement over the 1994 SREP submittals (65%).

PROGRAM EVALUATION: Each laboratory focal point evaluated ten areas (see Appendix 2) with a rating from one (lowest) to five (highest). The distribution of ratings was as follows:



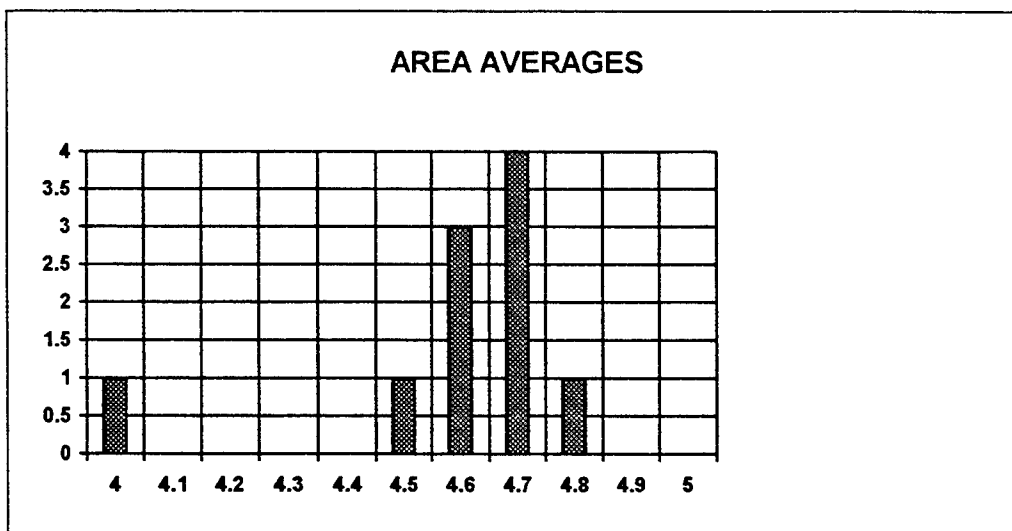
Rating	Not Rated	1	2	3	4	5
# Responses	7	1	7	62 (6%)	226 (25%)	617 (67%)

The 8 low ratings (one 1 and seven 2's) were for question 5 (one 2) "The USAF should continue to pursue the research in this SREP report" and question 10 (one 1 and six 2's) "The

one-year period for complete SREP research is about right”, in addition over 30% of the threes (20 of 62) were for question ten. The average rating by question was:

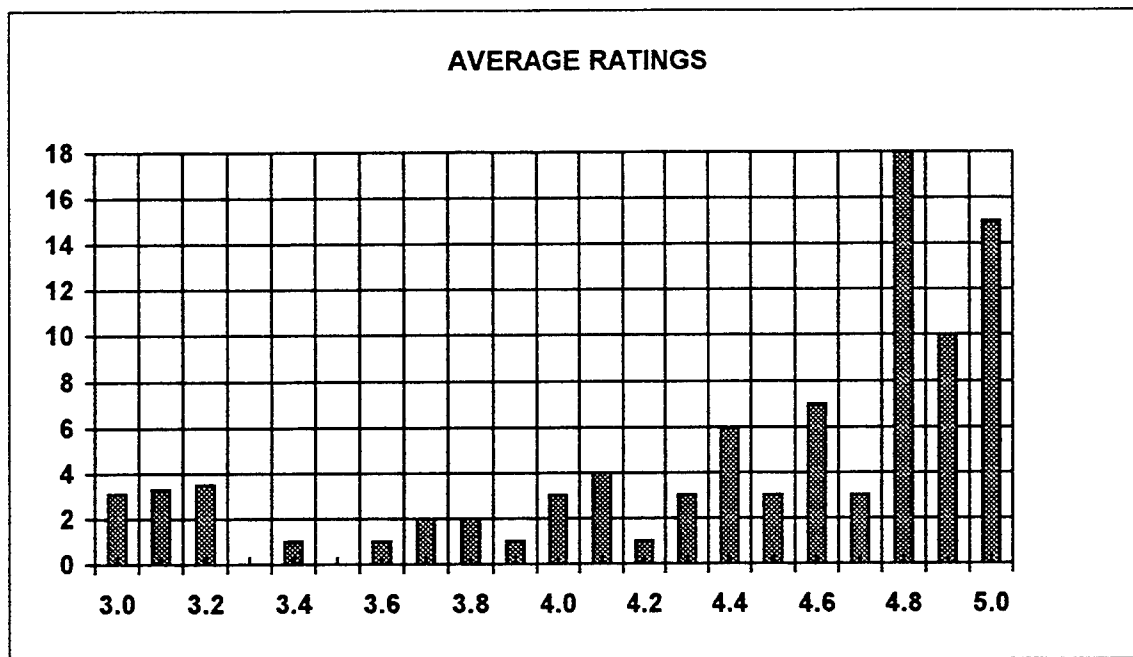
Question	1	2	3	4	5	6	7	8	9	10
Average	4.6	4.6	4.7	4.7	4.6	4.7	4.8	4.5	4.6	4.0

The distribution of the averages was:



Area 10 “the one-year period for complete SREP research is about right” had the lowest average rating (4.1). The overall average across all factors was 4.6 with a small sample standard deviation of 0.2. The average rating for area 10 (4.1) is approximately three sigma lower than the overall average (4.6) indicating that a significant number of the evaluators feel that a period of other than one year should be available for complete SREP research.

The average ratings ranged from 3.4 to 5.0. The overall average for those reports that were evaluated was 4.6. Since the distribution of the ratings is not a normal distribution the average of 4.6 is misleading. In fact over half of the reports received an average rating of 4.8 or higher. The distribution of the average report ratings is as shown:



It is clear from the high ratings that the laboratories place a high value on AFOSR's Summer Research Extension Programs.

3.0 SUBCONTRACTS SUMMARY

Table 1 provides a summary of the SREP subcontracts. The individual reports are published in volumes as shown:

<u>Laboratory</u>	<u>Volume</u>
Armstrong Laboratory	1A, 1B
Arnold Engineering Development Center	5
Frank J. Seiler Research Laboratory	5
Phillips Laboratory	2
Rome Laboratory	3
Wilford Hall Medical Center	5
Wright Laboratory	4A, 4B

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Anderson , James Analytical Chemistry University of Georgia, Athens, GA	PhD 95-0807	AL/EQ	01/01/95 12/31/95 Determination of the Redox Capacity of Soil Sediment and Prediction of Pollutant	\$25000.00	\$1826.00
Ashrafiuon , Hashem Mechanical Engineering Villanova University, Villanova, PA	PhD 95-0800	AL/CF	01/01/95 12/31/95 Finite Element Modeling of the Human Neck and Its Validation for the ATB Model	\$25000.00	\$19528.00
Burke , Michael Tulane University Tulane University, New Orleans, LA	PhD 95-0811	AL/HR	01/01/95 09/30/95 An Examination of the Validity of the New Air Force ASVAB Composites	\$25000.00	\$1818.00
Edwards , Paul Chemistry Edinboro Univ of Pennsylvania, Edinboro, PA	PhD 95-0808	AL/EQ	01/01/95 12/31/95 Fuel Identification by Neural Networks Analysis of the Response of Vapor Sensiti	\$25000.00	\$5000.00
Gerstman , Bernard Physics Florida International Universi, Miami, FL	PhD 95-0815	AL/OE	01/01/95 12/31/95 A Comparison of Multistep vs Singlestep Arrhenius Integral Models for Describing	\$24289.00	\$2874.00
Graetz , Kenneth Department of Psychology University of Dayton, Dayton, OH	PhD 95-0812	AL/HR	01/01/95 12/31/95 Effects of Mental Workload and Electronic Support on Negotiation Performance	\$25000.00	\$0.00
Gupta , Pushpa Mathematics University of Maine, Orono, ME	PhD 95-0802	AL/AO	01/01/95 12/31/95 Regression to the Mean in Half Life Studies	\$25000.00	\$2859.00
Koch , Manfred Geophysics Florida State University, Tallahassee, FL	PhD 95-0809	AL/EQ	12/01/94 04/30/95 Application of the MT3D Solute Transport Model to the Made-2 Site: Calibration	\$25000.00	\$0.00
Novotny , Mark Supercomputer Comp Res. I Florida State University, Tallahassee, FL	PhD 95-0810	AL/EQ	01/01/95 12/31/95 Computer Calculations of Gas-Phase Reaction Rate Constants	\$25000.00	\$0.00
Nurre , Joseph Mechanical Engineering Ohio University, Athens, OH	PhD 95-0804	AL/CF	01/01/95 12/31/95 Surface Fitting Three Dimensional Human Head Scan Data	\$25000.00	\$20550.00
Piepmeier , Edward Pharmaceutics University of South Carolina, Columbia, SC	PhD 95-0801	AL/AO	01/01/95 12/31/95 The Effects of Hyperbaric Oxygenation on Metabolism of Drugs and Other Xenobioti	\$25000.00	\$11740.00
Quinones , Miguel Psychology Rice University, Houston, TX	PhD 95-0813	AL/HR	01/01/95 12/31/95 Maintaining Skills After Training: The Role of Opportunity to Perform Trained T	\$25000.00	\$4000.00
Riccio , Gary Psychology Univ of IL Urbana-Champaign, Urbana, IL	PhD 95-0806	AL/CF	01/01/95 05/31/95 Nonlinear Transcutaneous Electrical Stimulation of the Vestibular System	\$22931.00	\$0.00
Shebilske , Wayne Dept of Psychology Texas A&M University, College Station, TX	PhD 95-0814	AL/HR	01/01/95 12/31/95 Cognitive Factors in Distr Training Effects During Acquisition of Complex Skills	\$25000.00	\$5614.00

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Weisenberger , Janet Dept of Speech & Hearing Ohio State University, Columbus, OH	PhD 95-0805	AL/CF	01/01/95 12/31/95 Tactile Feedback for Simulation of Object Shape and Textural Information in Hapt	\$25000.00	\$12234.00
Hughes , Rod Psychology Oregon Health Sciences University, Portland, OR	MA 95-0803	AL/CF	01/01/95 12/31/95 Melatonin Induced Prophylactic Sleep as a Countermeasure for Sleep Deprivation	\$25000.00	\$0.00
Bapty , Theodore Electrical Engineering Vanderbilt University, Nashville, TN	MS 95-0848	AEDC/E	01/01/95 12/31/95 Plant-Wide Preventive Maintenance & Monitoring	\$24979.00	\$0.00
Dorgan , John Chemical Engineering Colorado School of Mines, Golden, CO	PhD 95-0834	FJSRL/F	01/01/95 12/31/95 Block Copolymers at Inorganic Solid Surfaces	\$25000.00	\$0.00
Jungbauer , Mary Ann Chemistry Barry University, Miami, FL	PhD 95-0836	FJSRL/F	01/01/95 12/31/95 Non-Linear Optical Properties of Polyacetylenes and Related Substituted Compound	\$25000.00	\$24714.00
Statman , David Physics Allegheny College, Meadville, PA	PhD 95-0835	FJSRL/F	01/01/95 12/31/95 Studies of Second Harmonic Generation in Glass Waveguides	\$25000.00	\$6500.00
, Krishnaswamy Aeronautics University of Houston, Houston, TX	PhD 95-0818	PL/RK	01/01/95 12/31/95 Mixed-Mode Fracture of Solid Propellants	\$24993.00	\$8969.00
Ashgriz , Nasser Mechanical Engineering SUNY-Buffalo, Buffalo, NY	PhD 95-0816	PL/RK	01/01/95 12/31/95 Effects of the Jet Characteristics on the Atomization and Mixing in A Pair of Im	\$25000.00	\$22329.00
Bellem , Raymond Computer Science Embry-Riddle Aeronautical Univ, Prescott, AZ	PhD 95-0817	PL/VT	12/01/94 11/30/95 Experimental Studies of the Effects of Ionizing Radiation on Commerically Proces	\$20000.00	\$8293.00
Brzosko , Jan Nuclear Physics Stevens Institute of Tech, Hoboken, NJ	PhD 95-0828	PL/WS	11/01/94 02/01/95 Neutron Diagnostics for Pulsed Plasmas of Compact Toroid - Marauder Type	\$24943.00	\$0.00
Damodaran , Meledath Math & Computer Science University of Houston-Victoria, Victoria, TX	PhD 95-0831	PL/LI	01/01/95 12/31/94 Parallel Computation of Zernike Aberration Coefficients for Optical Aber Correct	\$24989.00	\$9850.00
DeLyser , Ronald Electrical Engineering University of Denver, Denver, CO	PhD 95-0877	PL/WS	01/01/95 12/31/95 Quality Factor Evaluation of Complex Cavities	\$25000.00	\$46066.00
Diels , Jean-Claude Physics University of New Mexico, Albuquerque, NM	PhD 95-0819	PL/LI	01/01/95 12/31/95 Unidirectional Ring Lasers and Laseer Gyros with Multiple Quantum Well Gain Medi	\$25000.00	\$0.00
Henson , James Electrical Engineering University of Nevada, Reno, NV	PhD 95-0820	PL/WS	01/01/95 12/31/95 Automatic Feature Extraction and Assessment of Wideband Range-Doppler Imagery of	\$25000.00	\$0.00
Kaiser , Gerald Physics University of Mass/Lowell, Lowell, MA	PhD 95-0821	PL/GP	01/01/95 12/31/95 Multiresolution Analysis with Physical Wavelets	\$25000.00	\$5041.00

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Kowalak , Albert Chemistry University of Massachusetts/Lo, Lowell, MA	PhD 95-0822	PL/GP The Synthesis and Chemistry of Peroxonitrites and Peroxonitrous Acid	01/01/95 12/31/95	\$24996.00	\$4038.00
Malloy , Kevin Electrical Engineering University of New Mexico, Albuquerque, NM	PhD 95-0829	PL/VT Temperature & Pressure Dependence of the Band Gaps & Band Offsets	01/01/95 12/31/95	\$24999.00	\$0.00
Prasad , Sudhakar Physics University of New Mexico, Albuquerque, NM	PhD 95-0823	PL/LI Theoretical Studies of the Performance of Novel Fiber-Coupled Imaging Interferom	01/01/95 12/31/95	\$25000.00	\$11047.00
Purtill , Mark Mathematics Texas A&M Univ-Kingsville, Kingsville, TX	PhD 95-0824	PL/WS Static and Dynamic Graph Embedding for Parallel Programming	01/01/95 12/31/95	\$25000.00	\$100.00
Rudolph , Wolfgang Physics University of New Mexico, Albuquerque, NM	PhD 95-0833	PL/LI Ultrafast Process and Modulation in Iodine Lasers	01/01/95 12/31/95	\$24982.00	\$6000.00
Stone , Alexander Mathematics & Statistics University of New Mexico, Alburquerque, NM	PhD 95-0827	PL/WS Impedance Matching And Reflection Minimization For Transient EM Pulses Through D	01/01/95 12/31/95	\$24969.00	\$0.00
Swenson , Charles Dept of Electrical Engr Utah State University, Logan, UT	PhD 95-0826	PL/VT Low Power Retromodulator based Optical Transceiver for Satellite Communications	01/01/95 12/31/95	\$25000.00	\$25000.00
Lipp , John Electrical Engineering Michigan Technological Univ, Houghton, MI	MS 95-0832	PL/LI Improved Methods of Tilt Measurement for Extended Images in the Presence of Atmo	01/01/95 12/31/95	\$24340.00	\$15200.00
Petroski , Janet Chemistry Cal State Univ/Northridge, Northridge, CA	BA 95-0830	PL/RK Thermoluminescence of Simple Species in Molecular Hydrogen Matrices	10/01/94 12/31/94	\$4279.00	\$0.00
Salasovich , Richard Mechanical Engineering University of Cincinnati, Cincinnati, OH	MS 95-0825	PL/VT Design, Fabrication, Intelligent Cure, Testing, and Flight Qualification of an A	01/01/95 12/31/95	\$25000.00	\$4094.00
Aalo , Valentine Dept of Electrical Engr Florida Atlantic University, Boca Raton, FL	PhD 95-0837	RL/C3 Performance Study of an ATM/Satellite Network	01/01/95 12/31/95	\$25000.00	\$13120.00
Amin , Moeness Electrical Engineering Villanova University, Villanova, PA	PhD 95-0838	RL/C3 Interference Excision in Spread Spectrum Communication Systems Using Time-Freque	01/01/95 12/31/95	\$25000.00	\$34000.00
Benjamin , David Computer Science Oklahoma State University, Stillwater, OK	PhD 95-0839	RL/C3 Designing Software by Decomposition using KIDS	01/01/95 12/31/95	\$24970.00	\$0.00
Choudhury , Ajit Engineering Howard University, Washington, DC	PhD 95-0840	RL/OC Detection Performance of Over Resolved Targets with Non-Uniform and Non-Gaussian	11/30/94 10/31/95	\$25000.00	\$0.00
Harackiewicz , Frances Electrical Engineering So. Illinois Univ-Carbondale, Carbondale, IL	PhD 95-0841	RL/ER Computer-Aided-Design Program for Solderless Coupling Between Microstrip and Str	01/01/95 12/31/95	\$23750.00	\$29372.00

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Losiewicz , Beth Psycholinguistics Colorado State University, Fort Collins, CO	PhD 95-0842	RL/IR	01/01/95 12/31/95 Spanish Dialect Identification Project	\$25000.00	\$4850.00
Musavi , Mohamad University of Maine, Orono, ME	PhD 95-0843	RL/IR	01/01/95 12/31/95 Automatic Image Registration Using Digital Terrain Elevation Data	\$25000.00	\$12473.00
Norgard , John Elec & Comp Engineering Univ of Colorado-Colorado Sprg, Colorado	PhD 95-0844	RL/ER	01/01/95 12/31/95 Infrared Images of Electromagnetic Fields	\$25000.00	\$2500.00
Richardson , Dean Photonics SUNY Institute of Technology, Utica, NY	PhD 95-0845	RL/OC	01/01/95 12/31/95 Femtosecond Pump-Probe Spectroscopy System	\$25000.00	\$15000.00
Ryder, Jr. , Daniel Chemical Engineering Tufts University, Medford, MA	PhD 95-0846	RL/ER	01/01/95 12/31/95 Synthesis and Properties B-Diketonate-Modified Heterobimetallic Alkoxides	\$25000.00	\$0.00
Zhang , Xi-Cheng Physics Rensselaer Polytechnic Institu, Troy, NY	PhD 95-0847	RL/ER	01/01/95 12/31/95 Optoelectronic Study of Seniconductor Surfaces and Interfaces	\$25000.00	\$0.00
Drost-Hansen , Walter Chemistry University of Miami, Coral Gables, FL	PhD 95-0875	WHMC/	01/01/95 12/31/95 Biochemical & Cell Physiological Aspects of Hyperthermia	\$25000.00	\$8525.00
Baginski , Michael Electrical Engineering Auburn University, Auburn, AL	PhD 95-0869	WL/MN	01/01/95 12/31/95 An Investigation of the Heating and Temperature Distribution in Electrically Exc	\$24995.00	\$10098.00
Berdichevsky , Victor Aerospace Engineering Wayne State University, Detroit, MI	PhD 95-0849	WL/FI	01/01/95 12/31/95 Micromechanics of Creep in Metals and Ceramics at High Temperature	\$25000.00	\$0.00
Buckner , Steven Chemistry Colullmbus College, Columbus, GA	PhD 95-0850	WL/PO	01/01/95 12/31/95 Development of a Fluorescenece-Based Chemical Sensor for Simultaneous Oxygen Qua	\$24900.00	\$8500.00
Carroll , James Electrical Engineering Clarkson University, Potsdam, NY	PhD 95-0881	WL/PO	01/01/95 12/31/95 Development of HIGH-Performance Active Dynamometer System for Machines and Drive	\$24944.00	\$38964.00
Choate , David Mathematics Transylvania University, Lexington, KY	PhD 95-0851	WL/AA	01/01/95 12/31/95 SOLVING $z(t)=\ln\{1/A[\cos(w_1t)]+B[\sin(w_2t)]+C\}$	\$24993.00	\$8637.00
Clarson , Stephen Materials Sci & Eng University of Cincinnati, Cincinnati, OH	PhD 95-0852	WL/ML	12/01/94 11/30/95 Synthesis, Processing and Characterization of Nonlinear Optical Polymer Thin Fil	\$25000.00	\$15000.00
Cone , Milton Comp Science & Elec Eng Embry-Riddel Aeronautical Univ, Prescott, AZ	PhD 95-0853	WL/AA	01/01/95 12/31/95 An Investigation of Planning and Scheduling Algorithms for Sensor Management	\$25000.00	\$11247.00
Courter , Robert Mechanical Engineering Louisiana State University, Baton Rouge, LA	PhD 95-0854	WL/MN	01/01/95 12/31/95 A Study to Determine Wave Gun Firing Cycles for High Performance Model Launches	\$25000.00	\$3729.00

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount	Univ. Cost Share
Dominic , Vincent Electro Optics Program University of Dayton, Dayton, OH	PhD 95-0868	WL/ML	01/01/95	12/31/95	\$25000.00	\$12029.00
		Characterization of Electro-Optic Polymers				
Fadel , Georges Dept of Mechanical Engr Clemson University, Clemson, SC	PhD 95-0855	WL/MT	01/01/95	12/31/95	\$25000.00	\$8645.00
		A Methodology for Affordability in the Design Process				
Gould , Richard Mechanical Engineering North Carolina State Univ, Raleigh, NC	PhD 95-0856	WL/PO	01/01/95	12/31/95	\$24998.00	\$9783.00
		Data Reduction and Analysis for laser Doppler Velocimetry				
Hardie , Russell Electrical Engineering Univsty of Dayton, Dayton, OH	PhD 95-0882	WL/AA	01/01/95	12/31/95	\$24999.00	\$7415.00
		Hyperspectral Target Identification Using Bomen Spectrometer Data				
Hodel , Alan Electrical Engineering Auburn University, Auburn, AL	PhD 95-0870	WL/MN	01/01/95	12/31/95	\$24990.00	\$9291.00
		Robust Falut Tolerant Control: Fault Detection and Classification				
Janus , Jonathan Aerospace Engineering Mississippi State University, Mississippi State,	PhD 95-0871	WL/MN	01/01/95	12/31/95	\$25000.00	\$7143.00
		Multidimensional Algorithm Development & Analysis				
Jasiuk , Iwona Dept of Materials Science Michigan State University, East Lansing, MI	PhD 95-0857	WL/ML	01/01/95	12/31/95	\$25000.00	\$0.00
		Characterization of Interfaces in Metal-Matrix Composites				
Jouny , Ismail Electrical Engineering Lafayette College, Easton, PA	PhD 95-0880	WL/AA	01/01/95	12/31/95	\$24300.00	\$5200.00
		TSI Mitigation: A Mountaintop Database Study				
Li , Jian Electrical Engineering University of Florida, Gainesville, FL	PhD 95-0859	WL/AA	10/10/95	12/31/95	\$25000.00	\$4000.00
		Comparative Study and Performance Analysis of High Resolution SAR Imaging Techni				
Lin , Chun-Shin Electrical Engineering University of Missouri-Columbi, Columbia, MO	PhD 95-0883	WL/FI	01/01/95	12/31/95	\$25000.00	\$2057.00
		Prediction of Missile Trajectory				
Lin , Paul Mechanical Engineering Cleveland State University, Cleveland, OH	PhD 95-0860	WL/FI	01/01/95	12/31/95	\$25000.00	\$6886.00
		Three Dimensional Deformation Comparison Between Bias and Radial Aircraft Tires				
Liou , Juin Electrical Engineering University of Central Florida, Orlando, FL	PhD 95-0876	WL/EL	01/01/95	12/31/95	\$25000.00	\$11040.00
		Investigation of AlGaAs/GaAs Heterojunction Bipolar Transister Reliability Based				
Nandhakumar , Nagaraj Electrical Engineering University of Virginia, Charlottesville, VA	PhD 95-0861	WL/AA	01/01/95	12/31/95	\$24979.00	\$4500.00
		Thermophysical Invariants fro, LWIR Imagery for ATR				
Pasala , Krishna Dept of Electrical Engr University of Dayton, Dayton, OH	PhD 95-0879	WL/AA	01/01/95	12/31/95	\$25000.00	\$1078.00
		Effect of Electromagmetic Enviornment on Array Signal Processing				
Perkowski , Marek Dept of Electrical Engr Portland State University, Portland, OR	PhD 95-0878	WL/AA	01/01/95	09/15/95	\$24947.00	\$18319.00
		Functional Decomposition of Binary, Multiple-Valued, & Fuzzy Logic				

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount	Univ. Cost Share
Reeves , Stanley Dept of Electrical Engnr Auburn University, Auburn, AL	PhD 95-0862	WL/MN	01/01/95	12/31/95	\$25000.00	\$0.00
		Superresolution of Passive Millimeter-Wave Imaging				
Rule , William Engineering Mechanics University of Alabama, Tuscaloosa, AL	PhD 95-0872	WL/MN	01/01/95	12/31/95	\$24968.00	\$14576.00
		Development of a Penetrator Optimizer				
Schauer , John Mech & Aerosp Eng University of Dayton, Dayton, OH	PhD 95-0873	WL/PO	11/01/94	11/30/95	\$25000.00	\$7428.00
		Heat Transfer for Turbine Blade Film Cooling with Free Stream Turbulence - Measu				
Schwartz , Carla Electrical Engineering University of Florida, Gainesville, FL	PhD 95-0863	WL/FI	01/01/95	12/31/95	\$25000.00	\$0.00
		Neural Network Identification and Control in Metal Forging				
Simon , Terrence Dept of Mechanical Engineering University of Minnesota, Minneapolis, MN	PhD 95-0864	WL/PO	01/01/95	12/31/95	\$24966.00	\$3996.00
		Documentation of Separating and Separated Boundary Layer Flow, for Application				
Skowronski , Marek Solid State Physics Carnegie Melon University, Pittsburgh, PA	PhD 95-0865	WL/EL	01/01/95	12/31/95	\$25000.00	\$6829.00
		Transmission Electron Microscopy of Semiconductor Heterojunctions				
Thirunarayan , Krishnaprasad Computer Science Wright State University, Dayton, OH	PhD 95-0866	WL/EL	01/01/95	12/31/95	\$25000.00	\$2816.00
		VHDL-93 Parser in SWI-PROLOG: A Basis for Design Query System				
Trelease , Robert Dept of Anatomy & Cell Bi University of California, Los Angeles, CA	PhD 95-0867	WL/ML	12/01/94	12/01/95	\$25000.00	\$0.00
		Development of Qualitative Process Control Discovery Systems for Polymar Composi				
Tsai , Chi-Tay Engineering Mechanics Florida Atlantic University, Boca Raton, FL	PhD 95-0874	WL/MN	01/01/95	12/31/95	\$24980.00	\$0.00
		Improved Algorithm Development of Massively Parallel Epic Hydrocode in Cray T3D				
Lewis , John Materials Science Engrng University of Kentucky, Lexington, KY	MS 95-0858	WL/MN	01/01/95	12/31/95	\$25000.00	\$13833.00
		The Characterization of the Mechanical Properties of Materials in a Biaxial Stre				

APPENDIX 1:
SAMPLE SREP SUBCONTRACT

**AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1995 SUMMER RESEARCH EXTENSION PROGRAM
SUBCONTRACT 95-0837**

BETWEEN

Research & Development Laboratories
5800 Uplander Way
Culver City, CA 90230-6608

AND

Florida Atlantic University
Department of Electrical Engineering
Boca Raton, FL 33431

REFERENCE: Summer Research Extension Program Proposal 95-0837
Start Date: 01-01-95 End Date: 12-31-95
Proposal Amount: \$25,000.00

- (1) **PRINCIPAL INVESTIGATOR:** Dr. Valentine A. Aalo
Department of Electrical Engineering
Florida Atlantic University
Boca Raton, FL 33431
- (2) **UNITED STATES AFOSR CONTRACT NUMBER:** F49620-93-C-0063
- (3) **CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER (CFDA):**12.800
PROJECT TITLE: AIR FORCE DEFENSE RESEARCH SOURCES PROGRAM
- (4) **ATTACHMENT 1 REPORT OF INVENTIONS AND SUBCONTRACT**
2 CONTRACT CLAUSES
3 FINAL REPORT INSTRUCTIONS

*****SIGN SREP SUBCONTRACT AND RETURN TO RDL*****

1. BACKGROUND: Research & Development Laboratories (RDL) is under contract (F49620-93-C-0063) to the United States Air Force to administer the Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), Bolling Air Force Base, D.C. Under the SRP, a selected number of college faculty members and graduate students spend part of the summer conducting research in Air Force laboratories. After completion of the summer tour participants may submit, through their home institutions, proposals for follow-on research. The follow-on research is known as the Summer Research Extension Program (SREP). Approximately 61 SREP proposals annually will be selected by the Air Force for funding of up to \$25,000; shared funding by the academic institution is encouraged. SREP efforts selected for funding are administered by RDL through subcontracts with the institutions. This subcontract represents an agreement between RDL and the institution herein designated in Section 5 below.

2. RDL PAYMENTS: RDL will provide the following payments to SREP institutions:

- 80 percent of the negotiated SREP dollar amount at the start of the SREP research period.
- The remainder of the funds within 30 days after receipt at RDL of the acceptable written final report for the SREP research.

3. INSTITUTION'S RESPONSIBILITIES: As a subcontractor to RDL, the institution designated on the title page will:

- a. Assure that the research performed and the resources utilized adhere to those defined in the SREP proposal.
- b. Provide the level and amounts of institutional support specified in the SREP proposal..
- c. Notify RDL as soon as possible, but not later than 30 days, of any changes in 3a or 3b above, or any change to the assignment or amount of participation of the Principal Investigator designated on the title page.
- d. Assure that the research is completed and the final report is delivered to RDL not later than twelve months from the effective date of this subcontract, but no later than December 31, 1998. The effective date of the subcontract is one week after the date that the institution's contracting representative signs this subcontract, but no later than January 15, 1998.
- e. Assure that the final report is submitted in accordance with Attachment 3.
- f. Agree that any release of information relating to this subcontract (news releases, articles, manuscripts, brochures, advertisements, still and motion pictures, speeches, trade associations meetings, symposia, etc.) will include a statement that the project or effort depicted was or is sponsored by: Air Force Office of Scientific Research, Bolling AFB, D.C.
- g. Notify RDL of inventions or patents claimed as the result of this research as specified in Attachment 1.
- h. RDL is required by the prime contract to flow down patent rights and technical data requirements to this subcontract. Attachment 2 to this subcontract

contains a list of contract clauses incorporated by reference in the prime contract.

4. All notices to RDL shall be addressed to:

RDL AFOSR Program Office
5800 Uplander Way
Culver City, CA 90230-6609

5. By their signatures below, the parties agree to provisions of this subcontract.



Abe Sopher
RDL Contracts Manager

Signature of Institution Contracting Official

Typed/Printed Name

Date

Title

Institution

Date/Phone

ATTACHMENT 2
CONTRACT CLAUSES

This contract incorporates by reference the following clauses of the Federal Acquisition Regulations (FAR), with the same force and effect as if they were given in full text. Upon request, the Contracting Officer or RDL will make their full text available (FAR 52.252-2).

FAR CLAUSES

TITLE AND DATE

52.202-1	DEFINITIONS
52.203-3	GRATUITIES
52.203-5	COVENANT AGAINST CONTINGENT FEES
52.203-6	RESTRICTIONS ON SUBCONTRACTOR SALES TO THE GOVERNMENT
52.203-7	ANTI-KICKBACK PROCEDURES
52.203-8	CANCELLATION, RECISSION, AND RECOVERY OF FUNDS FOR ILLEGAL OR IMPROPER ACTIVITY
52.203-10	PRICE OR FEE ADJUSTMENT FOR ILLEGAL OR IMPROPER ACTIVITY
52.203-12	LIMITATION ON PAYMENTS TO INFLUENCE CERTAIN FEDERAL TRANSACTIONS
52.204-2	SECURITY REQUIREMENTS
52.209-6	PROTECTING THE GOVERNMENT'S INTEREST WHEN SUBCONTRACTING WITH CONTRACTORS DEBARRED, SUSPENDED, OR PROPOSED FOR DEBARMENT
52.212-8	DEFENSE PRIORITY AND ALLOCATION REQUIREMENTS
52.215-2	AUDIT AND RECORDS - NEGOTIATION
52.215-10	PRICE REDUCTION FOR DEFECTIVE COST OR PRICING DATA

52.215-12	SUBCONTRACTOR COST OR PRICING DATA
52.215-14	INTEGRITY OF UNIT PRICES
52.215-8	ORDER OF PRECEDENCE
52.215.18	REVERSION OR ADJUSTMENT OF PLANS FOR POSTRETIREMENT BENEFITS OTHER THAN PENSIONS
52.222-3	CONVICT LABOR
52.222-26	EQUAL OPPORTUNITY
52.222-35	AFFIRMATIVE ACTION FOR SPECIAL DISABLED AND VIETNAM ERA VETERANS
52.222-36	AFFIRMATIVE ACTION FOR HANDICAPPED WORKERS
52.222-37	EMPLOYMENT REPORTS ON SPECIAL DISABLED VETERAN AND VETERANS OF THE VIETNAM ERA
52.223-2	CLEAN AIR AND WATER
52.223-6	DRUG-FREE WORKPLACE
52.224-1	PRIVACY ACT NOTIFICATION
52.224-2	PRIVACY ACT
52.225-13	RESTRICTIONS ON CONTRACTING WITH SANCTIONED PERSONS
52.227-1	ALT. I - AUTHORIZATION AND CONSENT
52.227-2	NOTICE AND ASSISTANCE REGARDING PATIENT AND COPYRIGHT INFRINGEMENT

52.227-10	FILING OF PATENT APPLICATIONS - CLASSIFIED SUBJECT MATTER
52.227-11	PATENT RIGHTS - RETENTION BY THE CONTRACTOR (SHORT FORM)
52.228-7	INSURANCE - LIABILITY TO THIRD PERSONS
52.230-5	COST ACCOUNTING STANDARDS - EDUCATIONAL INSTRUCTIONS
52.232-23	ALT. I - ASSIGNMENT OF CLAIMS
52.233-1	DISPUTES
52.233-3	ALT. I - PROTEST AFTER AWARD
52.237-3	CONTINUITY OF SERVICES
52.246-25	LIMITATION OF LIABILITY - SERVICES
52.247-63	PREFERENCE FOR U.S. - FLAG AIR CARRIERS
52.249-5	TERMINATION FOR CONVENIENCE OF THE GOVERNMENT (EDUCATIONAL AND OTHER NONPROFIT INSTITUTIONS)
52.249-14	EXCUSABLE DELAYS
52.251-1	GOVERNMENT SUPPLY SOURCES

DOD FAR CLAUSES**DESCRIPTION**

252.203-7001	SPECIAL PROHIBITION ON EMPLOYMENT
252.215-7000	PRICING ADJUSTMENTS
252.233-7004	DRUG FREE WORKPLACE (APPLIES TO SUBCONTRACTS WHERE THERE IS ACCESS TO CLASSIFIED INFORMATION)
252.225-7001	BUY AMERICAN ACT AND BALANCE OF PAYMENTS PROGRAM
252.225-7002	QUALIFYING COUNTRY SOURCES AS SUBCONTRACTS
252.227-7013	RIGHTS IN TECHNICAL DATA - NONCOMMERCIAL ITEMS
252.227-7030	TECHNICAL DATA - WITHOLDING PAYMENT
252.227-7037	VALIDATION OF RESTRICTIVE MARKINGS ON TECHNICAL DATA
252.231-7000	SUPPLEMENTAL COST PRINCIPLES
252.232-7006	REDUCTIONS OR SUSPENSION OF CONTRACT PAYMENTS UPON FINDING OF FRAUD

APPENDIX 2:

SAMPLE TECHNICAL EVALUATION FORM

SUMMER RESEARCH EXTENSION PROGRAM TECHNICAL EVALUATION

SREP NO: 95-0811

SREP PRINCIPAL INVESTIGATOR: Dr. Michael Burke

Circle the rating level number, 1 (low) through 5 (high), you feel best evaluate each statement and return the completed form by mail to:

RDL
Attn: 1995 SREP Tech Evals
5800 Uplander Way
Culver City, CA 90230-6608
(310) 216-5940 or (800) 677-1363

-
- | | | |
|-----|---|-----------|
| 1. | This SREP report has a high level of technical merit. | 1 2 3 4 5 |
| 2. | The SREP program is important to accomplishing the lab's mission. | 1 2 3 4 5 |
| 3. | This SREP report accomplished what the associate's proposal promised. | 1 2 3 4 5 |
| 4. | This SREP report addresses area(s) important to the USAF. | 1 2 3 4 5 |
| 5. | The USAF should continue to pursue the research in this SREP report. | 1 2 3 4 5 |
| 6. | The USAF should maintain research relationships with this SREP associate. | 1 2 3 4 5 |
| 7. | The money spent on this SREP effort was well worth it. | 1 2 3 4 5 |
| 8. | This SREP report is well organized and well written. | 1 2 3 4 5 |
| 9. | I'll be eager to be a focal point for summer and SREP associates in the future. | 1 2 3 4 5 |
| 10. | The one-year period for complete SREP research is about right. | 1 2 3 4 5 |
-

11. If you could change any one thing about the SREP program, what would you change.

12. What would you definitely NOT change about the SREP program?

USE THE BACK FOR ANY ADDITIONAL COMMENTS.

Laboratory: Armstrong Laboratory
Lab Focal Point: Linda Sawin Office Symbol: AL/HRMI
Phone: (210) 536-3876

INVESTIGATION OF THE MIXED-MODE FRACTURE BEHAVIOR OF SOLID PROPELLANTS

Krishnaswamy Ravi-Chandar
Associate Professor
Department of Mechanical Engineering

University of Houston
Houston, TX 77204-4792

Final Report for:
Summer Research Extension Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Phillips Laboratory

December 1995

INVESTIGATION OF THE MIXED-MODE FRACTURE BEHAVIOR OF SOLID PROPELLANTS

Krishnaswamy Ravi-Chandar
Associate Professor
Department of Mechanical Engineering
University of Houston

Abstract

Experiments were performed to examine the mixed-mode fracture behavior of solid propellants. In these experiments, the deformation field near the crack tip was monitored using moire interferometry, providing a quantitative characterization of the displacements. The displacement fields were compared to that predicted through linear elastic fracture mechanics. The stress intensity factors were then determined through a least square fitting of the observed and calculated moire fringe patterns. It was determined by a comparison of the measured and the predicted crack opening profile that a linear elastic fracture mechanics approach would be adequate in these experiments. The resulting fracture behavior was compared with the predictions of the maximum tangential stress theory. While this theory appears to predict the failure load levels adequately, it does not predict the kinking angle correctly. It is suggested that damage development at the crack tip and rate effects play a significant role in mixed-mode fracture and are probably responsible for this.

INVESTIGATION OF THE MIXED-MODE FRACTURE BEHAVIOR OF SOLID PROPELLANTS

K. Ravi-Chandar

Introduction

In an effort to examine the mixed-mode fracture of solid propellants, preliminary tasks were undertaken earlier to develop the tools necessary for such an investigation. After investigating the methods of caustics, speckle interferometry and moire interferometry, it was decided that moire interferometry would be the most easily adapted technique for use in solid propellants (Ravi-Chandar, 1995). In this report, we present the results of some experiments aimed at investigating the mixed-mode fracture of solid propellants. We describe first the theoretical formulations typically used in mixed-mode fracture analysis, then the technique of moire interferometry, followed by the experimental scheme used to perform the experiments; finally the results are interpreted in terms of the theoretical formulations along with a description of future work needed to resolve some of the issues raised.

Mixed-Mode Fracture Characterization

The mixed-mode problem considered here involves determination of critical conditions at which a crack subjected to a combination of loading modes I and II will initiate. Another important consideration is the crack trajectory upon initiation. Since the early work of Erdogan and Sih (1962), this problem has received considerable attention (for example, Williams and Ewing, (1972); Maiti and Smith (1983), Ueda et al (1983)) and a number of criteria have been proposed to predict crack initiation behavior under mixed-mode loading. In the two-dimensional case, the Cartesian components of the crack tip stresses can be expressed (in terms of the polar coordinates as)

$$\sigma_{\alpha\beta}(r, \theta) = \frac{K_I}{\sqrt{2\pi r}} f_{\alpha\beta}(\theta) + \frac{K_{II}}{\sqrt{2\pi r}} g_{\alpha\beta}(\theta); r \rightarrow 0, \alpha, \beta = 1, 2, \quad (1)$$

where $f_{\alpha\beta}(\theta)$ and $g_{\alpha\beta}(\theta)$ are characteristic functions and K_I and K_{II} are the mode I and mode II stress intensity factors respectively. The asymmetry of the loading is generally described in terms of the mixity parameter μ which is defined as

$$\mu = \frac{K_{II}}{K_I}. \quad (2)$$

Thus the mixed-mode fracture problem involves determining critical stress intensity factor pairs (K_I^c, K_{II}^c) for different values of μ , at which the crack will initiate. Furthermore it is required to determine the initiation angle γ as a function of μ , as well as the subsequent crack propagation path. Two different criteria that have been proposed to predict the critical condition for crack initiation and the crack trajectory upon initiation are reviewed here.

The Maximum Tangential Stress Criterion (MTS): This criterion was proposed by Erdogan and Sih (1962) and it is stated as follows: Crack extension starts at the crack tip in a radial direction. *This extension is in that radial direction perpendicular to the direction of the greatest tension. Crack extension begins when this tension reaches a certain critical value at a certain distance from the crack tip.* Erdogan and Sih used a plate containing a central crack of length $2a$ subjected to uniaxial far-field loading σ , the crack being oriented at angle β to the loading direction. They obtained a relation for the initiation angle γ as a function of μ by equating the partial derivative $\partial\sigma_{\theta\theta} / \partial\theta$ to zero (equivalently the shear stress $\sigma_{r\theta}$ can be equated to zero). Both these conditions yield the relation

$$\sin \gamma + (3 \cos \gamma - 1)\mu = 0. \quad (3)$$

From eq.(3) it can be shown that for pure mode II conditions ($\mu = 0$) the initiation angle is predicted to be 70.5° .

The Strain Energy Density Criterion (SED): This theory was proposed by Sih (1973) and it involves the following two hypotheses: *The crack extends in a radial direction that is characterized by the minimum strain energy density. Crack extension begins when the strain energy density in this particular direction reaches a critical value.* The strain energy density is defined in terms of K_I and K_{II} as

$$S = a_{11}K_I^2 + a_{12}K_I K_{II} + a_{22}K_{II}^2 \quad (4)$$

where a_{11} , a_{12} and a_{22} are functions of θ and are defined according to the conditions of plane stress and plane strain, in terms of the Poisson's ratio. The results from most experiments were more in agreement with the MTS criterion than with the SED criterion. Apart from the above mentioned, four more criteria viz the maximum potential energy release rate criterion, the Maximum Tangential Strain criterion (MTSN) based on the near tip strains, the Energy Momentum Tensor criterion based on Eshelby's energy momentum vector, and the T-criterion have been proposed by Palaniswamy and Knauss (1978), Wu (1978), Tirosh (1977) and Theocaris (1982) respectively. However, in accordance with the trend shown in literature, attention is focussed on the first two criteria.

Dependence on the Rate of Loading: In the case of mixed mode loading, the cross-head speed is not representative of the local rate of loading at the crack tip. The appropriate rate of loading, if the MTS

criterion is to be applied, is the rate of increase of $\sigma_{\theta\theta}$ along the direction γ of maximum tangential stress. This rate, $\partial\sigma_{\theta\theta}(\gamma)/\partial t$, depends not only on the cross-head rate, but also on the mixity parameter μ and decreases as μ increases. Thus, for mixed-mode experiments in rate dependent materials, rate effects are combined with load asymmetry. Cotterell and Rice (1980) show that the local stress intensity factors k_1 and k_2 at the tip of an infinitesimal crack at an angle α with the main crack are given (to first order) by

$$\begin{aligned} k_1 &= \frac{K_I}{4} \left[3 \cos \frac{\alpha}{2} + \cos \frac{3\alpha}{2} - 3\mu \left(\sin \frac{\alpha}{2} + \sin \frac{3\alpha}{2} \right) \right] \\ k_2 &= \frac{K_{II}}{4} \left[3 \sin \frac{\alpha}{2} + \sin \frac{3\alpha}{2} + 3\mu \left(\cos \frac{\alpha}{2} + \cos \frac{3\alpha}{2} \right) \right] \end{aligned} \quad (5)$$

where K_I and K_{II} ($=\mu K_I$) are the stress intensity factors at the main crack prior to kinking. The impact of rate of loading can be made clear using equations (5). Under mixed-mode loading conditions, the cross-head speed determines the rate of increase of K_I and K_{II} , but the rate of increase of k_1 at the incipient kink is dependent not only on K_I but also μ . Equations (5) imply that for a fixed cross-head rate, dk_1/dt decreases as μ increases. Thus in materials with rate dependent fracture properties, such as solid propellants, the MTS criterion must be used with appropriate consideration of the local rate of loading; an example of this was demonstrated for a thermoset polyester, Homalite-100, by Mahajan and Ravi-Chandar (1989).

Experimental Arrangement For Mixed-Mode Loading

We examined the use of a compact-tension-shear specimen to investigate the mixed mode fracture criteria for solid propellants. This specimen, shown in Figure 1 along with the loading grips, was introduced by Buchholz *et al* (22) and has been successfully in our laboratory in the past in examining mixed mode fracture in polymers (Mahajan and Ravi-Chandar, (1989)) and polymer interface cracks (Foltyn and Ravi-Chandar, (1990)). The stress intensity factors K_I and K_{II} for this geometry are given as (see Buchholz *et al* (22)):

$$\begin{aligned} \frac{K_I}{\frac{P}{wt}\sqrt{\pi a}} &= \frac{\cos \alpha}{1 - \frac{a}{w}} \sqrt{\frac{0.26 + 2.65\left(\frac{a}{w-a}\right)}{1 + 0.55\left(\frac{a}{w-a}\right) - 0.08\left(\frac{a}{w-a}\right)^2}} \\ \frac{K_{II}}{\frac{P}{wt}\sqrt{\pi a}} &= \frac{\sin \alpha}{1 - \frac{a}{w}} \sqrt{\frac{-0.23 + 1.40\left(\frac{a}{w-a}\right)}{1 - 0.67\left(\frac{a}{a-w}\right) + 2.08\left(\frac{a}{a-w}\right)^2}} \end{aligned} \quad (6)$$

where α is the angle between the normal to the crack and the load line, P is the applied load, a is the crack length, w is the specimen width and t is the specimen thickness. The crack was introduced into the specimen by cutting a slit to nearly half the width of the specimen. The specimen and loading attachments were placed on an Instron test frame; the displacement field near the crack tip was monitored using moire interferometry as described in the next section. The experiments were performed under a constant rate extension. Since the conditions local to the crack tip were measured by the moire, crack tip conditions can be measured quantitatively even after large crack extension, without the need for extensive computations. Figure 2 shows the variation of the load-load-point displacement for the eight tests that were performed. A consistent trend that is observed is that the peak load at fracture increases as the mode II component of the loading is increased. We will return to a consideration of these results after describing the technique of moire interferometry and the interpretation of the crack tip displacement fields.

Moire Interferometry

Post et al (1988) had developed a grid deposition technique for propellants: The specimen surface was coated with a grid of 100 lines per inch using a wire grid mask and dusting fine particles of TiO_2 . Smith, Rezvani and Liu (1989) had used this technique to determine the displacements and strains by photographing and analyzing the grids before and after deformation. In some earlier work (Ravi-Chandar, 1995), we changed this technique into direct moire interferometry by placing a glass master grating of 100 lines per inch to view the resulting moire fringes, corresponding to the displacement component normal to the crack line. This technique worked well and was used to examine the crack tip deformation fields under mixed-mode loading. Eight samples were examined; in this report, the moire interferometry results from two of the samples are discussed. A 4"x2"x0.5" specimen, with a single edge crack was loaded in pure mode I, and another similar specimen was loaded in a combined mode I and mode II configuration. A field of view of 2"x2" centered at the crack tip was used to observe the moire fringes of the crack normal displacement component. The resulting moire fringe pattern was recorded on a video system for later analysis.

Results

One frame from the the video sequence of a mode I test is shown in Figure 3a. These fringes correspond to contours of constant displacement in the crack normal direction. This data was analysed by obtaining the displacements at different distances from the crack tip and comparing them to the theoretically expected displacement if a K-field were appropriate. Thus, the experimental data set of $V^E(x,y)$ was compared to $V^{Th}(x,y;K_I,K_{II})$ using a least square error χ^2

$$\chi^2 = \sum_{i=1}^N \left[\frac{V^E(x_i, y_i) - V^{Th}(x_i, y_i; K_I, K_{II})}{\sigma} \right]^2 \quad (7)$$

in order to determine K_I and K_{II} . The crack opening displacement was then calculated and compared to the measured crack opening; from the best fit values of K_I and K_{II} moire fringes were also simulated and compared to the experimental fringes. In Figure 3b, the experimentally observed moire data are compared with the simulation from the best fit; also shown in this figure is a comparison of the crack opening displacement obtained from the video and the best fit calculation. The moire fringe comparison indicates that a K-field might be appropriate in front of the crack in a sector $-\pi/4 < \theta < \pi/4$, $3\text{mm} < r < 5\text{mm}$. While there is a large discrepancy between the simulated and measured fringes outside of this area, the calculated crack opening displacement - calculated based on a K-field curve fit to the experimental data in the sector indicated above - compared very well with the measured crack opening displacement profile even 15 mm behind the crack tip!

A similar investigation was performed for the case of mixed-mode loading; figure 4a shows one frame from the video sequence of the mixed mode test. The moire interference fringe patterns again correspond to contours of constant displacement in the direction normal to the initial crack. Once again by using the least square fit idea of equation (7), the mixed-mode stress intensity factors K_I and K_{II} were determined. The crack opening displacement and moire fringe patterns were then simulated with the best fit values of K_I and K_{II} and are compared in figure 4b; once again the comparison indicates that in the region of around 3 to 5 mm from the crack tip, in a sector ahead of the crack tip, the K-field is considered to be appropriate. Also, the crack opening displacement appears to agree quite well for a much larger range. In conclusion, these tests indicate that a K-field approach might be quite meaningful for this solid propellant, if applied after careful evaluation of the region in which the dominance of the K-field is prevalent. We now turn to an examination of the mixed-mode fracture results.

The critical values of the stress intensity factors at the onset of failure were computed using Eq.(6) and are plotted in Figure 5 as the failure interaction diagram. The prediction of the maximum tangential stress theory is also shown in this figure; the strain energy density theory does not differ significantly from this and hence is not indicated in the plot. Figure 6 shows the variation of the kink angle with the mode mix, characterized here by the parameter $\delta = \mu/(1+\mu)$. $\delta = 0$ corresponds to pure mode I and $\delta = 1$ corresponds to pure mode II. Clearly, the MTS theory does not predict the kink direction, even when the mode mix is only slightly different from a pure mode I loading. It appears that, for small amounts of mode II, the kink angle is equal to the loading angle, suggesting that the crack growth direction is normal to the loading direction. This is clearly seen in figure 7, where the kinking angle is plotted against the loading angle. At very large amounts of mode II, this idea does not appear to be appropriate. It appears that the maximum tangential stress criterion is not appropriate for the solid propellant; while its predictions for the stress intensity factor pairs at failure are not far from the experimental data, the kink angle predictions are quite different. Two causes might contribute to this: first, the damage that occurs in the neighborhood of the crack tip might

locally alter the stress field and thus influence the kinking angle. Secondly, propellants exhibit significant viscoelasticity and hence rate effects could be very important in determining the fracture characteristics. These issues need further investigation.

Conclusion

This report describes the results of an investigation into the mixed-mode fracture of solid propellants. Using the compact tension-shear specimen and moire interferometry, it has been demonstrated that a large range of mixed-mode loading can be generated in propellant specimens and that the use of the idea of a crack tip governed by a stress intensity factor is quite appropriate. It was found, however, that the maximum tangential stress criterion which is most commonly used in mixed-mode fracture characterization of other materials does not predict the kinking angle well in propellants, perhaps due to the influence of the damage near the crack tip and the rate dependence of the material.

References

- F.G. Buchholz, P.J.M. Pirro, H.A. Richard and K.H. Dreyer, 1987, Proceedings, Fourth International Conference on Numerical Methods in Fracture Mechanics, (eds Luxmore et.al), Pineridge Press, Swansea, UK, p.641.
- B. Cotterell and J.R. Rice, 1980, "Slightly curved or kinked cracks", International Journal of Fracture, 16, pp.155-169.
- F. Erdogan and G.C. Sih, 1963, "On the crack extension in plates under plane loading under transverse shear", Journal of Basic Engineering, 85D, pp.519-527.
- P. Foltyn and K. Ravi-Chandar, 1991, "Initiation of an interface crack under mixed-mode loading", accepted for publication in the Journal of Applied Mechanics.
- R.V. Mahajan and K. Ravi-Chandar, 1989, "An experimental investigation of mixed-mode fracture", International Journal of Fracture, 41, pp.235-252.
- S.K. Maiti and R.A. Smith, 1983, "Comparison of the criteria for mixed-mode brittle fracture based on the preinstability stress field. Part I: Slit and elliptical cracks under uniaxial tensile loading", International Journal of Fracture, 23, pp.281-295.
- K. Palaniswamy and W.G. Knauss, 1978, "On the problem of crack extension in brittle solids under general loading", Mechanics Today, Vol. 4, pp.87-148.
- D. Post, C.W. Smith and R. Czarnek, 1988, "Boundary layer and singularity measurement in three dimensional fracture problems", AFRPL-TR-88-086, Air Force Astronautics Laboratory, Edwards AFB, CA.
- K. Ravi-Chandar, 1995, "Experimental investigation of fracture in solid propellants", Report to Wyle Laboratories.
- G.C. Sih, 1973, "Some basic problems in fracture mechanics and new concepts", Engineering Fracture Mechanics, 5, p.365-377.

C.W. Smith, M. Rezvani and C.T. Liu, 1989, "Measurement of fracture parameters in polyphase materials", in *Micromechanics: Experimental Techniques*, ed. W.N. Sharpe, Jr., ASME AMD-Vol. 102, pp.107-119.

P.S. Theocaris, G. Kardomateas and N.P. Andrianopoulous, 1982, "Experimental study of the T-criterion in ductile fractures", *Engineering Fracture Mechanics*, 17, pp.439-447.

J. Tirosh, 1977, "Incipient fracture angle, fracture loci and critical stress for mixed-mode loading", *Engineering Fracture Mechanics*, 9, pp.607-616.

Y. Ueda et al, 1983, "Characteristics of brittle fracture under general combined modes including those under bi-axial tensile loads", *Engineering Fracture Mechanics*, 18, pp.1131-1158.

J.G. Williams and P.D. Ewing, 1972, "Fracture under complex stress- the angled crack problem", *International Journal of Fracture*, 8, pp.441-446.

C.H. Wu, 1978, "Fracture under combined loads by maximum-energy-release rate criterion", *Journal of Applied Mechanics*, 45, pp.553-558.

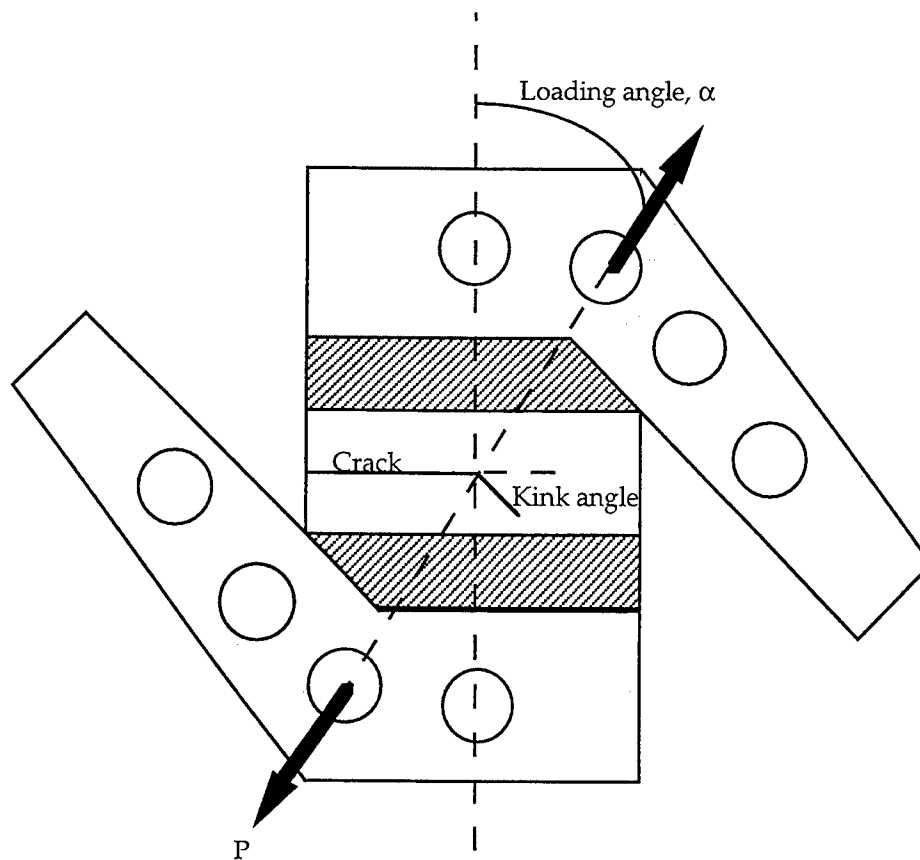


Figure 1. Geometry of the compact tension-shear specimen used for mixed-mode fracture studies.

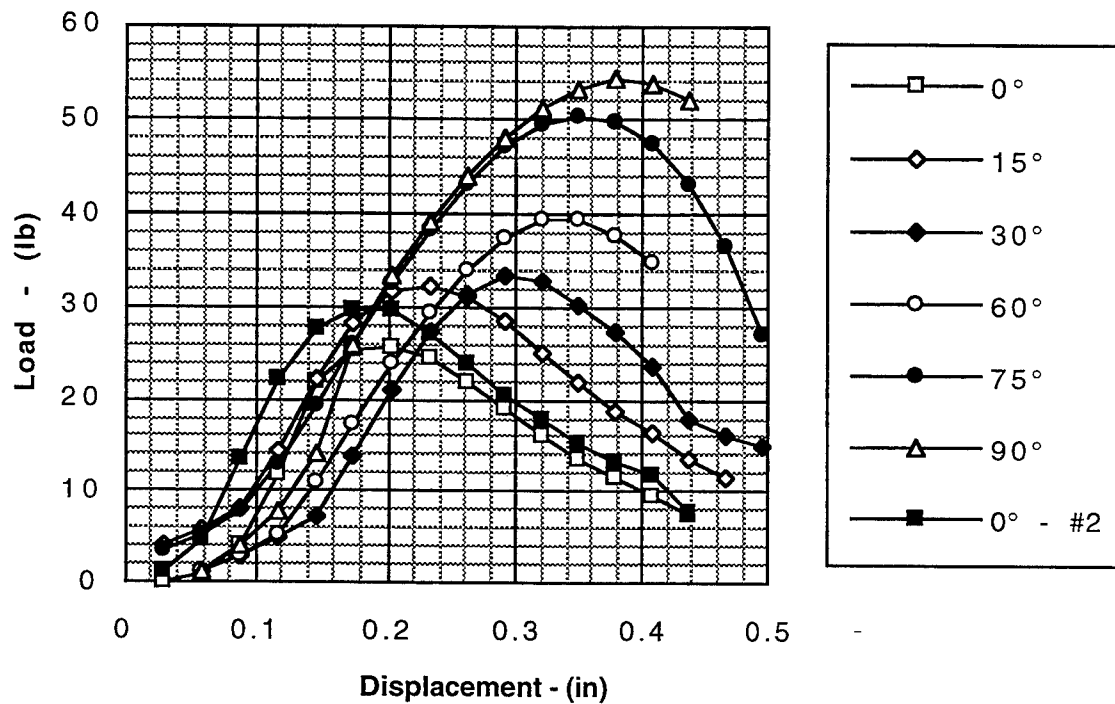
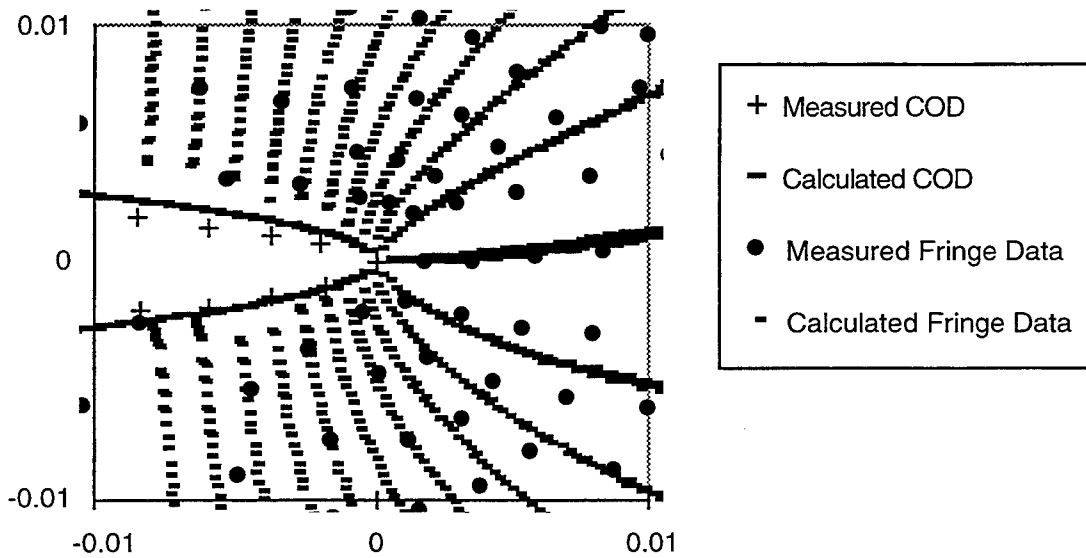


Figure 2. Load-load point displacement from mixed-mode experiments in solid propellant.



(a)



(b)

Figure 3. (a) Moire-interference fringe pattern near a crack tip corresponding to contours of constant crack normal displacement component; mode I loading. (b) Comparison of measured and calculated moire fringe patterns and crack opening displacement.



(a)

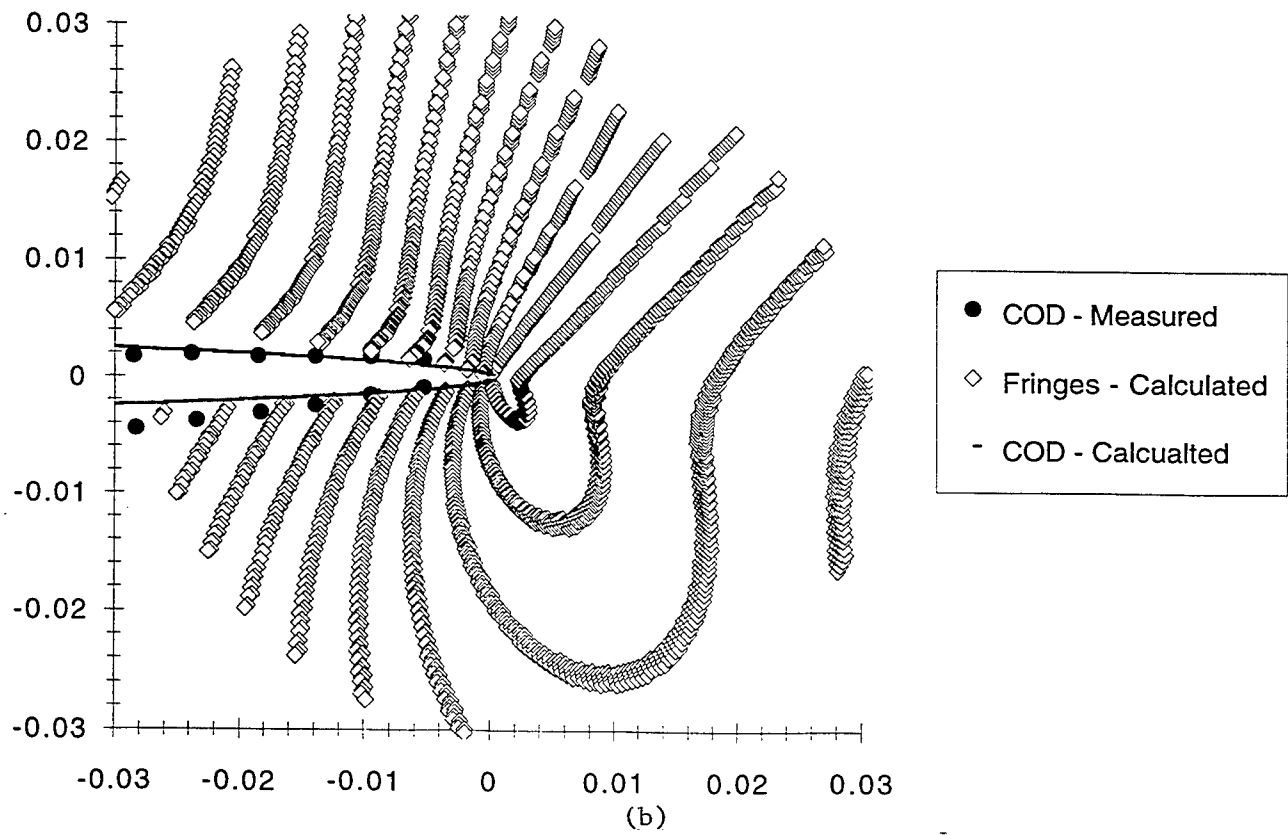


Figure 4. (a) Moire-interference fringe patterns near a crack tip corresponding to contours of constant crack normal displacement component; mixed-mode loading. (b) Comparison of measured and calculated moire fringe patterns and crack opening displacement.

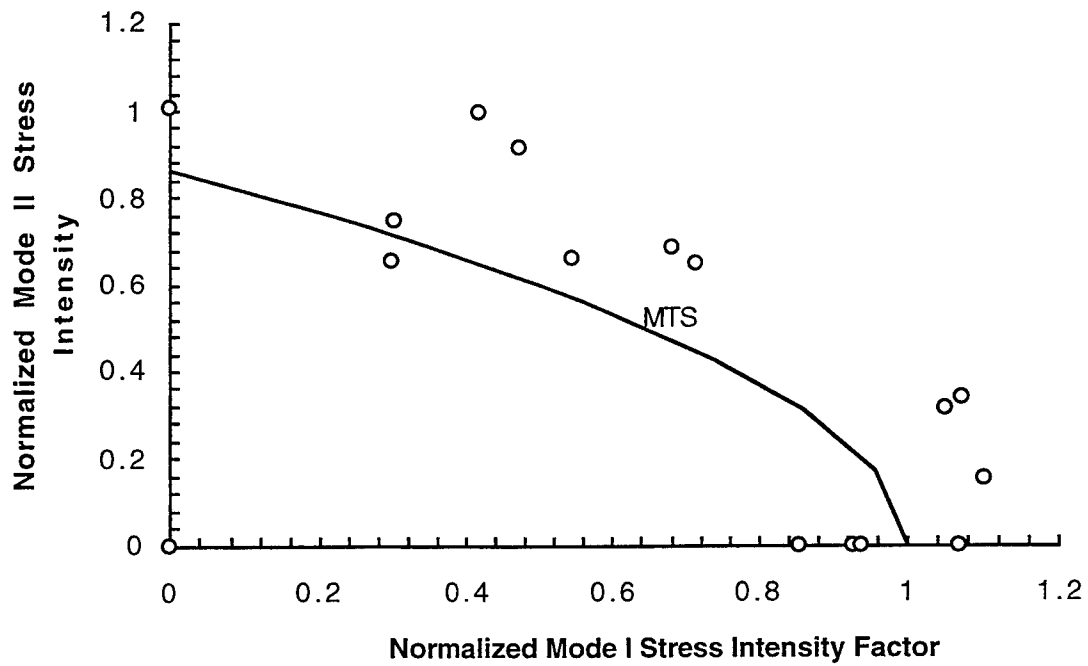


Figure 5. Mixed-mode failure interaction curve.

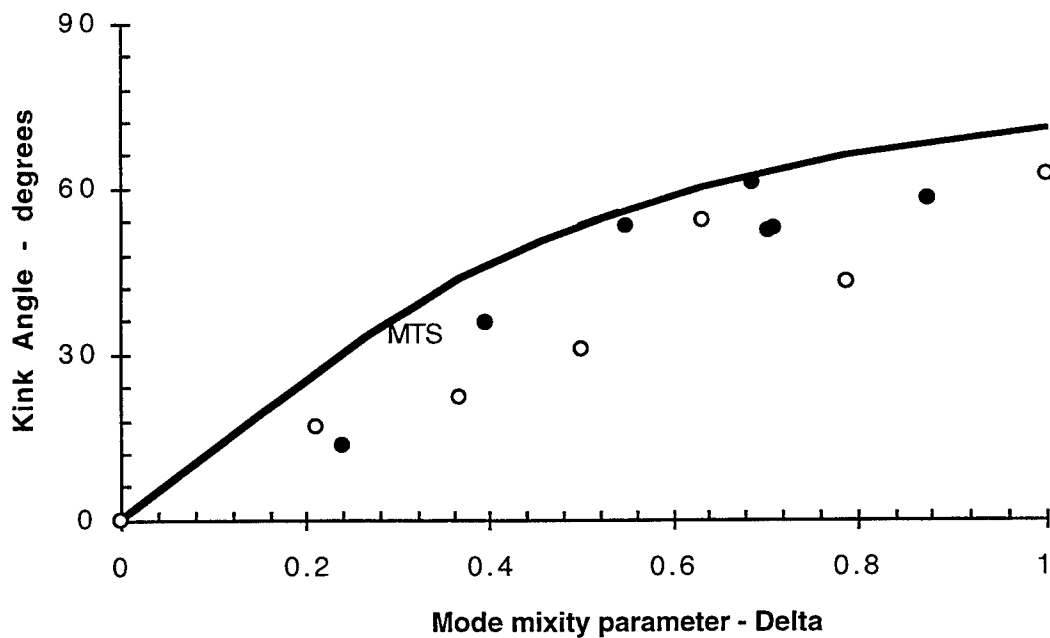


Figure 6. Dependence of the kink angle on the mode mix parameter δ .

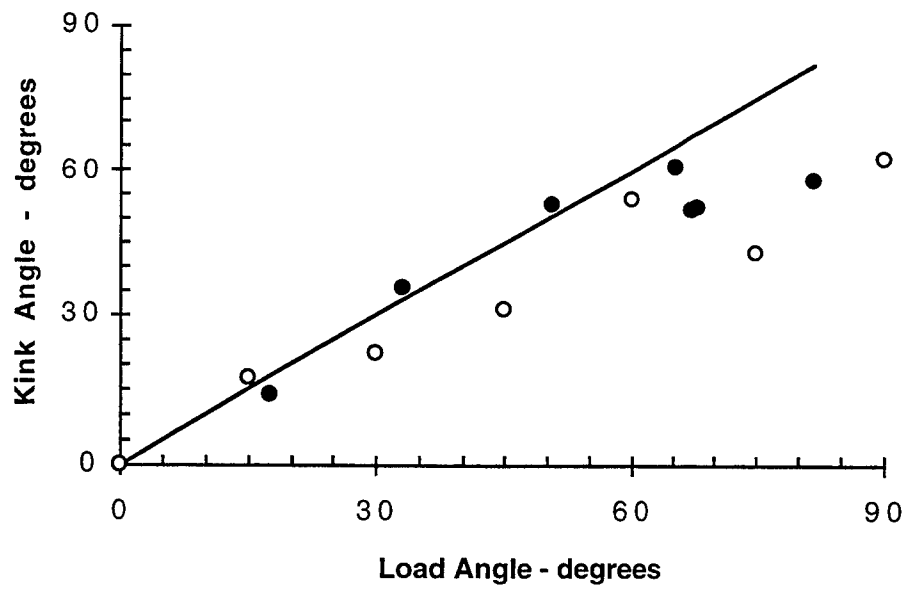


Figure 7. Dependence of the kink angle on the loading angle.

Performance Study of ATM-Satellite Network

Nasser Ashgriz
Associate Professor
Department of Mechanical and Aerospace Engineering

State University of New York at Buffalo
Buffalo, NY

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

Phillips Laboratory

December 1995

CHARACTERIZATION OF A TURBULENT LIQUID JET CLOSE TO THE ORIFICE

Nasser Ashgriz
Department of Mechanical and
Aerospace Engineering
State University of New York at Buffalo
Buffalo, New York 14260
e-mail: ashgriz@eng.buffalo.edu

ABSTRACT

Experimental investigation of mixing in a pair of impinging jets has revealed that at low jet Reynolds numbers the two jets seem to bounce off each other, whereas, at high jet Reynolds numbers they seem to cross through each other. The reasons for the latter case is not clear. It has been postulated that the helical structures of the jet prior to impingement are responsible for this crossing. In this project we have investigated the structure of the jets prior to impingement (a short distance from the nozzle) to identify the existence of the helical structures and their possible influence on the jet mixing. The results show that the helical structures exist for relatively low Reynolds number jets and they are mainly a downstream feature - it takes many jet diameters for these structures to form. For high Reynolds number jets the helical structures do not form prior to impingement. Therefore, the helical structures are not the cause of jet crossing in impinging jets. Rather, it is postulated that the initial momentum of each jet is the main cause of this transmissive mixing.

CHARACTERIZATION OF A TURBULENT LIQUID JET CLOSE TO THE ORIFICE

Nasser Ashgriz

1. Introduction

The performance of a biliquid propellant rocket motor is intimately dependent on the mixing and chemical reaction among the propellant or propellant combination in the combustor. In order to minimize the physical size of the combustion chamber, certain limitations are set on the residence time of the reactants in the chamber. These limitations dictate requirements for a rapid mixing and chemical reaction between the propellants. Such requirements can be achieved by proper atomization of the liquid propellants and attainment of a certain mass distribution of each propellant across the injection surface for a stoichiometric and/or optimum heat release rate. The process of mixing in actual combustors is governed by the simultaneous action of the vaporization, diffusion, and gaseous turbulence. However, the first order effect seems to be the original physical mixing or the mass distribution generated by the injector.

One of the techniques for the attainment of proper mixing is by impinging two or more liquid jets at a common point. In rocket engines, liquid jets of fuel and/or oxidizer collide and breakup into a spray that vaporizes and combusts. There are numerous experimental as well as theoretical studies on the mechanism of the atomization of the impinging jets. These studies, exemplified by the works of Heidmann and Humphrey (1951), Heidmann and Foster (1961), Rupe (1956), Taylor (1960), Dombrowski and Hooper (1963), Vassallo and Ashgriz (1992), and Anderson et al. (1992), show that when two jets collide they form a sheet in the direction perpendicular to the plane of jets. Waves form on the surface of the sheet and grow until the sheet breakups into small droplets. The generated spray has an elliptical cross-section with its major axis in the plane of the liquid sheet. A recent review by Ryan et al. (1993) addresses the various mechanisms which are believed to be responsible for the atomization of the impinging jets.

Studies on the mixing of the two streams are more scarce. There are several reports on the measurement of the spatial distribution of two components of a biliquid impinging stream spray such as those by Rupe (1953), Hoehn & Rupe (1971), Hoehn et al. (1972),

Ferrenberg & Jaqua (1983), Mchale & Nurick (1970, 1974), and Nurick (1969, 1976, 1991). The objectives have been to correlate spray properties, such as mixing efficiency, with the physical configuration of the injection system. The object of this work, however, is to understand the mechanism of mixing rather than determining the spray properties. Only recently, Ashgriz et al. (1995) provided some information on the mechanism of mixing in impinging-jet atomizers. A brief discussion of their findings is provided next.

2. Mixing Mechanisms

Ashgriz et al. (1995) measured the extent of mixing by adding a dye (food color) to one of the jets and collecting the liquid on the impingement plane (x -axis) using the linear Patternator. The concentration of the dye is then measured using a spectrophotometer. Detailed description of the experimental setup and other information of the jet characteristics are provided in Ashgriz et al. (1995) and will not be repeated here.

Figures 1-3 show the mixture fraction in the impingement plane for three different jet diameters, and several jet velocities. A very distinct feature is evident. At low jet velocities, when the two jets impinge and atomize, the liquid from each jet tends to stay in the same plane as the original jet. The two jets seem to bounce off each other. For instance, in Fig. 1-a ($D_j = 330\mu m$ and $V_j = 20m/s$) the jet on the left (negative x -axis) is dyed, and the mixture fraction, measured at $9.21cm$ downstream of the impingement point, continuously reduces from left to right. Similar results are obtained for jets with larger diameters, as shown in Fig. 2-a for $D_j = 508\mu m$ at $V_j = 13m/s$, and in Fig. 3-a for $D_j = 711\mu m$ at $V_j = 8m/s$. As the jet diameter is increased the bouncing type impingement occurs at lower jet velocities. When the jet velocity is increased, totally different distribution of the mixture fraction is obtained. The results indicate that the dyed jet crosses to the side of the undyed jet. This is evident from the plots of mixture fraction in Figure 1-b, 2-b, and 3-b,c, where the mixture fraction increases from left to right. Because of the turbulent nature of the flow, there is always some liquid from one jet which transmits to the other side. Therefore, we have not been able to find the limiting concentrations, i.e., zero and one, at the two extremum.

The observation that at low jet velocities, the two jets basically bounce off of each other after impingement, however, at high jet velocities they cross through each other, can be explained by dividing the mixing process into two different regions: Mixing in the liquid

sheet, i.e., pre-atomization zone; and mixing in the spray region, i.e., post-atomization zone.

If two liquid jets with equal diameters collide head-on, and the liquid in each stream is 'completely' deflected, two jets will not cross and a bouncing or 'reflecting type' atomization is obtained. A schematic representation of such impingement is shown in Fig. 4-a. The mixing is then governed by the turbulent dispersion in the spray zone. (The micromixing in the liquid sheet is relatively small.) On the other hand, one explanation for the observed stream crossing may be as follow. As the jet velocity increases, the sheet breaks up in a shorter distance from the impingement point and streams do not deflect completely. Therefore, one can assume a sheet with converging mean velocities, rather than parallel ones for the complete deflection. Figure. 4-c shows a schematic of this process. After atomization, the trajectory of each drop will be governed by the trajectory of the stream which has contributed the most to the liquid in that drop. Therefore, shortly after atomization, the drops segregate, and it appears that the two jets cross each other.

The above explanation of the mixing process is based on a head-on impingement of two jets. However, due to the turbulent nature of the streams, head-on impingement hardly ever occurs in a continuous fashion. In order to achieve proper atomization (relatively small drop sizes), the jets should have high velocities and turbulent conditions. Turbulent liquid jets have a very unique characteristics. They develop a helical disturbances due to the form drag, the amplitude of which is entirely a function of the relative velocity between the jet and the surrounding gas. Such helical disturbances have also been reported by the experiments of Hoyt and Taylor (1977) and predicted by the analytical studies of Batchelor and Gill (1962) and Yang (1992). The helical characteristics of the streams cause that the two jet do not collide head-on. Therefore, at the collision point part of the liquid of each stream which is not interacting with the other stream tends to follow its own initial trajectory. The consequence of this process is a stretching and oscillation of the liquid sheet shown in Fig. 5. Figure 6 shows the region of interaction between two jets when the collision is not head-on. The parameter which governs the extent of this interaction is the amplitude of the helical disturbances at the time of impingement.

The above description of the impinging jets, explain the experimental observations. Our results show that increasing the jet velocity and/or the impingement angle, enhances the liquid crossing. Since the region of each stream which does not interact with the oppos-

ing stream is assumed to be responsible for the crossing of the two stream, all the parameters which increase the horizontal component of the kinetic energy of this region (component perpendicular to the direction of the resultant velocity vector of the two stream) control the crossing process: $\frac{\pi}{8} \rho_l V_j^3 D_j^2 \sin(\alpha/2)$.

Both the reflective and transmissive types of impingements result in poor mixing. However, in the former the liquid from each stream stays on the its own side, yet in the latter it crosses to the other side. Therefore, at certain jet velocity optimum mixing can occur as depicted in Fig. 4-b. In other words the curves of mixing should show a maximum with jet velocity. However, at low jet velocities the spray dispersion in the plane of the jets (x -axis) was small and the resolution of our Patternator was not sufficient to capture small variations in the jet velocity. Therefore, the results of the reflective type impingement are not at sufficiently small velocities, and they are already close to the optimum mixing condition as shown in Fig. 1-a. Any increase from this condition reduces the mixing extent. Generally, higher jet velocities were needed for smaller jet diameters in order to obtain large enough dispersion in the plane of the jets for accurate measurement of the mixing.

The effect of increasing the impingement angle on the mixing is shown in Fig. 7. An increase in the impingement angle from 60° to 90° enhances the transmission, and the extent of mixing reduces. Rupe (1953) had explained this effect by noting that higher impingement angles result in a higher impact-induced turbulence and, therefore, higher turbulence mixing. On the other hand, higher impingement angles will result in smaller contact time between the liquids of the two streams. However, our results indicate that it is the higher momentum which enhances the stream crossing and results in poor mixing. As the impingement angle increases the component of the momentum in the horizontal direction increases, increasing the liquid crossing.

Plots of mixing extent Φ_m with the jet velocity, given in Fig. 8, show that the mixing reduces with the velocity. There is one data point for $D_j = 508\mu m$ which indicates that increase in V_j increases Φ_m . However, similar measurements closer to the injector (upstream) show that Φ_m decreases with increasing V_j . This result indicates that the mixing in the post-atomization region plays a significant role in the extent of mixing. Similar results are observed for a jet with $D_j = 711\mu m$.

The enhancement of mixing in the post-atomization region is better realized by plot-

ting the variation of Φ_m along the spray axis z , as shown in Fig. 9. In all the cases, except one, Φ_m increases downstream of the impingement point. If the jets are significantly segregated, by either crossing or reflecting type impingement, the segregation increases downstream of the spray. The turbulent dispersion may not be significant enough to compensate for this segregation and overall extent of mixing, Φ_m reduces. This effect is shown for the jet with $D_j = 508\mu m$ and $V_j = 13m/sec$. In the post-atomization region, the turbulent dispersion controls the mixing of the two streams. Therefore, a mixing layer is developed in the central region of the spray, which grows along the spray axis. This explains the observed improvement of the mixing downstream of the impingement point. A schematic of the enhancement of mixing in the post-atomization region is shown in Figs. 4-d and 4-e by representing the mixture fraction with colored drops.

3. Characterization of the Jet Prior to Impingement

Ashgriz et al. (1995) have proposed two models for the observed transmissive mixing in impinging jets. One is based on the helical structure of the jets prior to impingement, and the other is based on the directional momentum of the jets. In this study we have looked at the jet structures shortly after issuance from the orifice to determine the characteristics of the helical structures and address the above models of mixing.

Most investigations on the mixing problem have limited the jet characteristics to the following conditions: (1) The flow and the nozzle conditions are adjusted such that the jet diameter is approximately the same as the orifice diameter. In other words the jet does not spread significantly in the impingement distance. (2) The jet is kept stable both directionally and in time. Any unsteadiness in the flow velocity or direction can result in significant changes in the mixing and therefore, the analysis of the results becomes difficult.

In addition to the stream stability, another parameter which influences the mixing is the velocity profile in the jet before impingement. If the velocity profile of each of the streams is different from one another, a nonuniform momentum exchange and hence a non-symmetrical, distorted, and inconsistent spatial distribution will result. The impingement of such streams may be analogue to the impingement of idealized streams having different diameters, different momenta, and/or varying degree of misalignment. Rupe (1953) noted that turbulence-inducing devices tend to promote directional stability, reduce effects of upstream disturbance, and promote symmetry of the velocity profile.

Hoehn and Rupe (1971), and Nurick (1976) considered nozzles with different inlet conditions. They found that for circular unlike-impinging-doublet elements the mixing uniformity decreases with using sharp edged orifices. They attributed this reduction to the cavitation in the orifice. Orifices that were modified by length, entrance condition, and/or subjected to different operating conditions to prevent cavitation resulted in higher mixing factors. Noncavitating circular orifices produce maximum mixing factor of about 8% higher than those for the cavitating flow conditions.

In this project we have investigated the formation of the instabilities on the liquid jets and the influence of the jet velocity and the nozzle geometry on these instabilities.

3.1. Experimental Setup

The experimental setup consists of a flow stand, a traversing system, and the visualization system. A constant flow rate was essential to insure that the jet did not change while it was being characterized. The flow system consisted of a driving force, a holding tank, and gages. A high pressure tank of Nitrogen with a regulator attached was used to drive the flow. This delivered a constant pressure to a 30 liter stainless steel tanks filled with water. This type of flow system has the advantage to any water pump, since water supply provides a jet stream for the experiments free of pulsation and vibration. Reducing any pulsation and vibration of the flow is necessary to prevent triggering surface waves caused by flow disturbances upstream of the nozzle tip. The steady flow was monitored by an orifice flow meter and set with a needle valve. After setting the desired pressure drop across the orifice, the flow rate was calibrated with a graduated cylinder and stopwatch. The precision of the gage allowed the flow rate to be set to within 0.03ml/sec of the desired flow rate and as a result the jet velocity was determined to within 0.1m/sec .

The injectors were mounted onto a traversing stand that could run in the x , y , and z directions. The jet was moved from one position to another while the visualization system remained stationary. Since the characteristics of the jet change with position it is important to have the same reference point for each test. The origin of the coordinate axis for the experiment was the exit plane of the jet. The the x -axis is chosen to be the direction of the jet and the y -axis is taken to be perpendicular to the jet surface.

For this setup a total of 8 injectors was designed and built. The material used was 0.25 inch diameter stainless steal rod. A hole of 0.125 inch diameter was drilled to the desired

depth at which point the orifice diameter was drilled through the remaining material.

To visualize the surface waves of the turbulent jets, a Kodak Ektapro 1000 high speed camera was used. The recorded sessions are transferred to a VHS professional tape player and then projected onto a television for measurements. Once video clips of the jet surfaces were obtained, a frame grabber was used to take still photographs of the surfaces. These photos were then digitized and analyzed.

3.2. Fourier Analysis

In order to determine the dominant frequencies in the turbulent jet, the jet surface is decomposed into a set of harmonics. The surface function $f(x)$ may be written as

$$f(x) = \int_{-\infty}^{\infty} F(k)e^{ikx} dk$$

where

$$F(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x)e^{-ikx} dx$$

is the Fourier integral transform of $f(x)$. In the present study the variable x is the distance along the jet. Then k is the wavenumber or $2\pi/\lambda$ where λ is the wavelength. Since $e^{ikx} = \cos kx + i \sin kx$, the above equation provides a representation of the function $f(x)$ in terms of the periodic sine and cosine functions.

In our case the information on the location of the surface, or the height function $f(x)$, is provided in a discrete set. In such case one can write the Fourier integral transform for the band limited to $k < k_c$, where $k_c = 2\pi/\Delta x$ and Δx is the sample interval along the jet surface, as

$$F(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx e^{-ikx} \sum_{n=-\infty}^{\infty} F(n\Delta x) \text{sinc}(k_c x - n\pi)$$

$$F(k) = \sum_{n=-\infty}^{\infty} F(n\Delta x) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx} \text{sinc}(k_c x - n\pi)$$

where Shannon's sampling theorem has been used. Using the Fourier transform of the *sinc* function, the above equation becomes

$$F(k) = \begin{cases} \frac{\Delta x}{2\pi} \sum_{n=-\infty}^{\infty} F(n\Delta x) e^{-ikn\Delta x} & (k < k_c); \\ 0 & (\text{Otherwise}) \end{cases}$$

since $k_c = 2\pi/\Delta x$.

Since the amount of data is finite, measured values of $F(n\Delta x)$ for $n = 0, 1, 2, \dots, N-1$ are used. The finite Fourier transform then becomes:

$$F(k) \approx \frac{\Delta x}{2\pi} \sum_{n=0}^{N-1} F(n\Delta x) e^{-ikn\Delta x}$$

3.3. Results and Discussion

3.3.1. Flow Visualization

Typical pictures of the jet are shown in Figs. 10-12. Two distinct features are noted in these figures. For low Reynolds numbers, i.e., $Re=5000$ and 7000 , the jet surface contains relatively large wavelength disturbances. It appears as if there is a helical structure to the jet. From the instability theory the azimuthal disturbances form helical type of surface waves. The pictures show that the surface waves are not axisymmetric. Therefore, the observed waves should be generating azimuthal disturbances and hence helical structure. Helical structures had been observed by Roshko (1976) for an air jet but only far downstream from the nozzle. This was also predicted by Batchelor & Gill (1962). As the jet Reynolds number is increased helical type of structure are not observed any more (close to the orifice at least). The jet attains a very rough surface with many small perturbations. These disturbances are observed right at the exit of the orifice. Therefore, they are the consequence of the turbulent flow inside the injector. Probably, they represent the structure of the turbulent flow at the edges of the laminar boundary layer inside the injector. The pictures also show that the jet spread rate increases with increasing the Reynolds number. This indicates that the small surface perturbations grow due to the relative velocity between the jet and the ambient air. This growth has been shown to be due to the shear effects.

The main physical feature in liquid jets is the sudden removal of the wall shear-stress once the liquid stream separates from the solid injector. Near the point where the surface is

formed, at the large Reynolds numbers at least, the shear stress must fall from some finite value, at a point slightly within the liquid, to zero at the free surface. At the same time, the velocity of a fluid element on the surface increases from zero with distance downstream. It is shown by Goren (1966) that the surface velocity and surface position vary as the cube root of the distance downstream and with the shear rate of the flow. Therefore, in a turbulent stream if the shear rate fluctuates both in time and in position, similar fluctuations should be observed on the surface position both in time and in the azimuthal position on the surface of the jet. The small fluctuations observed on the jet are mainly due to these oscillations.

3.3.2. Various Modes of the surface disturbances

Figures 13-19 show the amplitude of each mode for different jets. In each figure five curves are provided. The curve marked with * represents the decomposition of a jet length which is 6.5 times orifice diameter ($L/D=6.5$). The other four curves provide the decomposition of a quarter of this length. This will allow us to determine how different modes evolve along the jet axis. Since the lengths are short, the results for the quarter of the jet length are less accurate than those for the whole jet. Therefore, we first identify the dominant modes on the whole jet and then use the data from the quarter of the length to postulate the evolution of each mode. Note that all the data sets contain amplitudes for $k = 0$. This amplitude is not relevant to our study. The reason it is not equal to zero is that the origin of the coordinate system is taken as the edge of the jet and not at the center of the wave train. In fact since the mean value of the fluctuations changes for each quadrant, the amplitude of the $k = 0$ changes with it. However, in the Fourier decomposition the amplitudes of various wavenumbers are independently determined. Therefore, the amplitudes of other wavenumbers can still be compared with each other even if we have not set the mean value of the fluctuations to zero.

Figure 13 shows the results for a jet with $Re = 5000$ issuing from an orifice with diameter $D_j = 1mm$ and $L_o/D_j = 3$, where L_o is the orifice length. The length of the jet which was characterized is $6.5mm$. The Fourier decomposition of the whole jet, $L/D = 6.5$, shows that the first 7 modes are all significant on the jet surface. In order to set a cut-off for the modes, we consider all the modes which have amplitudes larger than 0.03 as being significant. The odd modes, 1, 3, and 5 (but not the 7th mode) have slightly larger amplitudes than the even modes. Figure four indicates that disturbances with wavelengths

$\lambda_1 = 6.5$, $\lambda_2 = 3.25$, $\lambda_3 = 2.16$, $\lambda_4 = 1.625$, $\lambda_5 = 1.5$, $\lambda_6 = 1.08$, and $\lambda_7 = 0.93$ are observed on this jet. Figure 14 shows that modes 1, 3, and 6 are the dominant modes. However, the amplitudes in this case are lower than those in Fig. 13. The only difference between the jet in Fig. 13 and that in Fig. 14 is the orifice length. $(L/D)_o = 3$ in Fig. 13, whereas $(L/D)_o = 10$ in Fig. 14. The longer orifice allows the flow to become more developed and have less fluctuations. When the Reynolds number is increased to $Re = 18,000$, the only mode which has an amplitude larger than 0.03 is the second mode as shown in Fig. 15. However, a train of higher modes up to $m = 11$ are observed with small amplitudes. Figure 16 shows the modes for a jet with $Re=7000$ and $(L/D)_o = 5$. The jet diameter in this case is 0.6mm, which is smaller than that of previous figures. The dominant modes and their amplitudes are very close to those in Fig. 13, even though the jet diameter was almost halved. This indicates that the Reynolds number and the $(L/D)_o$ of the orifice are the controlling parameters for the jet surface disturbances. For this orifice when the Reynolds number is increased to $Re=18,000$, as shown in Fig. 17, the 1st mode becomes by far the dominant mode, having amplitudes of more than twice those of the next significant modes 2nd and 4th. Comparing these with those of Figure 16, it is clear that the increase in the Reynolds number has increased the amplitudes. This is in contrast with the jet which has $L/D = 10$ as shown in Figs. 14 and 15. It is well known that short orifices may cavitate at high Reynolds numbers. Therefore, for the short orifices the cavitation results in a larger fluctuations of the surface at higher Reynolds numbers and close to the nozzle. Figures 18 and 19 provide the surface modes for a jet with twice the orifice length as that of Figures 16 and 17. At the low Reynolds number of 7000, the results are almost the same as depicted in Figs. 16 and 18. The difference arises when the Reynolds number is increased to 18,000. It appears as if the longer length of the jet smooth out the velocity profile inside the injected after the formation of the cavitation. In Figure 19 the 1st mode has three times the amplitude of the next train of dominant modes.

In order to understand the significance of these modes a discussion of the theoretical analysis of the jet breakup is essential. Hoyt and Taylor(1977) describe the complex system of disturbances on the surface of a water jet discharging into air based on two loosely defined types, namely, "axially symmetric," and "helical" disturbances. The axially symmetric modes are rarely observed on the surface of a turbulent jet. Therefore, they are the sinuous waves which govern the breakup of the liquid jets. The three dimensional stability analyses of Yang (1992) and Li (1995) have shown that for a jet issuing in an

stagnant air the sinuous modes can be unstable if the Weber number is large enough. That is the cut-off wavenumber increases with the Weber number. The jets studies here have Weber numbers in the range of $1300 \leq We \leq 4300$. For this range of Weber numbers they predicted that the cutoff wavenumber for the first sinuous mode ($m=1$) will range between $0.6 \leq ka \leq 4$ where k is the wavenumber and it is related to the wavelength as $k = 2\pi/\lambda$ and a is the jet radius. For the higher modes, higher Weber numbers are needed to render the jet unstable. The cutoff wavenumbers of 0.6 and 4 result in wavelengths of 5.2mm and 0.785 for a jet with 1mm diameter and wavelengths of 3.14mm and 0.47mm for a jet with diameter of 0.6mm. Our experimental results show that the first modes for all our jets have wavelengths which are larger than the cutoff wavenumber. However, the higher modes become unstable only at higher Reynolds numbers or for smaller orifices. One reason that the amplitudes of the 1st and 2nd modes of the smaller diameter orifice was larger may be that the growth rates of the 1st and the 2nd modes are larger in these cases. Generally, the wavelengths of the dominant modes are in the same order of the unstable wavenumbers as predicted by the linear theory of Li (1995).

The genesis of these disturbances may be related to either the turbulent fluctuations inside the injector or the waves formed due to the transition region from laminar to turbulent. Based on the Mattingly & Chang's (1974) theory and using half the thickness of the laminar boundary layer, Hoyt and Taylor (1977) estimated a wavelength of .635mm for the waves accompanying transition. The wavelengths that we observe for the surface waves are about the same value only for mode 10, i.e., $\lambda_{10} = 5.6mm$. As it was noted these higher modes have very small amplitudes and are not dominating close to the injector. The wavelengths of corresponding to the turbulence fluctuations are even smaller. Therefore, the genesis of these disturbances must be related to some other larger scale of the flow fluctuations.

3.3.3. Spatial variation of the surface modes

Although as the length of the jet is reduces the accuracy of the determination of the surface modes reduces, once can deduce some general phenomenological information from such data. In this section we use the surface decomposition of a quarter of the jet length to obtain the surface modes.

In Fig. 13 the largest amplitude for the 1st quadrant occurs at $m = 3$. The largest amplitude for 2nd, 3rd, and 4th occur at $m = 1$. The length of the jet which is characterized

is $6.5mm$. therefore, the length of each quadrant is $1.625mm$. Then the wavelength for $m = 1$ is $\lambda_1 = 1.625mm$, for $m = 3$ is $\lambda_3 = 0.54mm$. Therefore, the results indicate that the smaller wavelengths are dominant close to the orifice, and the larger wavelengths dominate farther downstream of the jet. There seems to be a three fold increase in the wavelength between the two regions in the jet for this Reynolds number.

Figure 14 shows the amplitudes of various waves for a jet with $Re = 5000$ issuing from an orifice with diameter $D_j = 1mm$ and $L_o/D_j=10$. In this case $m = 1$ seems to be dominant in all quadrants. Therefore, the long length of the orifice eliminates the formation of the small wavelengths. This is may be due to the development of a laminar boundary layer inside the orifice which damps the formation of the high frequency surface disturbances due to the flow turbulence.

Figure 15 shows the results for the same nozzle as in Figure 14, but for a larger Reynolds number of $Re = 18,000$. For the 1st quadrant, no significant fluctuation is observed on the jet. Modes 1,2, and 3 all have a very small amplitudes. In the 2nd quadrant the maximum amplitude occurs at $m = 2$ ($\lambda_2 = 0.73mm$). in the 3rd quadrant it occurs at $m = 1$. and at the 4th quadrant it occurs at $m = 2$ again. Because of the high Reynolds number the disturbances appear further downstream of the nozzle. Similar results are obtained for Figs. 16-19.

4. Concluding Remarks

The results show that for low Reynolds number turbulent jets, the jet has a helical type of structure. The wavelength of these helical structures are about two times the jet diameter. However, the wavelength increases along the axis of the jet. The jet surface also contains small wavelength disturbance with about the same magnitude as the jet diameter. Formation of these helical disturbances are delayed to further downstream when the jet Reynolds number is increased. As the flow rates of the liquid is increased, the amplitudes of the surface waves dampen. Although, this is only applicable to the lower modes. For the higher modes the amplitudes at both Reynolds numbers are similar. As L/D is increased, its effect on the surface wave amplitude is to decrease it. The amplitudes are larger for the lower Reynolds numbers and small modes, which again shows that as the flow increased through the orifice the amplitudes of the surfaces become smaller. For a small L/D , the disturbance to the jet's internal flow is more turbulent allowing the creation of larger amplitude surface waves. It appears that the lower frequencies generated by the larger jet diameters grow at a higher rate than those for the smaller jet diameters.

For the conditions typically observed in impinging jet atomizers in rocket injectors, the pre-impingement length of the jet is short and the Reynolds number is high. Therefore, the helical structures do not have time to form and the phenomena is governed by relatively axisymmetric turbulent jet. It is therefore concluded that the transmissive mixing in impinging jet is not due to the helical disturbances. The other effect that remains is the directional momentum of the jets which is thought to be the cause of this stream crossing.

5. References

- Anderson, W.E., Ryan, H.M., Pal, S., and Santoro, R.J., "Fundamental studies of Impinging Liquid Jets," AIAA Paper No. 92-0458, 1992.
- Ashgriz, N., Brocklehurst, W., Talley, D., "On the Mixing Mechanisms in a Pair of Impinging Jets," AIAA Joint Propulsion Conference, July 10-12, 1995.
- Batchelor, G.K., & Gill, A.E., "Analysis of the Stability of Axisymmetric Jets," *J. Fluid Mech.*, 14, 529, 1962.
- Dombrowski, N., and Hooper, P.C., "A Study of the Sprays Formed by Impinging Jets in Laminar and Turbulent Flow," *J. Fluid Mech.*, 18, 1963.
- Ferrenberg, A., Jaqua, V., "Atomization and Mixing Study," Rockwell International, Rocketdyne Division, Report No. RI/RD83-170, 1983.
- Goren, S.L., "Development of the Boundary Layer at a Free Surface from a Uniform Shear Flow," *J. Fluid Mech.*, 25, 87-95, 1966.
- Heidmann, M.F., and Humphrey, J.C., "Fluctuations in a Spray Formed by two Impinging Jets", NACA Tech Note 2349, April 1951.
- Heidmann, M.F., and Foster, H.H., "Effect of Impingement Angle on Drop-Size Distribution and Spray Pattern of Two Impinging Water Jets", NASA Tech Note D- 872, July 1961.
- Hoehn, F., and Rupe, J., "Flow Cavitation Effects on the Mixing Characteristics of Bipropellant Doublets," 8th JANNAF Combustion Meeting, CPIA Pub. 220, Nov. 1971, Vol. 1, pp. 569-582.
- Hoehn, F.W., Rupe, J.H., and Sotter, J.G., "Liquid-Phase Mixing of Bipropellant Doublets," Jet Propulsion Laboratory, Technical Report 32-1546, 1972.
- Hoyt, J.W., and J.J. Taylor, "Waves on Water Jets," *J. Fluid Mech.* vol. 83, paper 1, pp. 119-127, 1977.
- Li, X., "Mechanism of Atomization of a Liquid Jet," *Atomization and Sprays*, Vol. 5, No. 1, pp. 89-106, 1995.
- Mattingly, G.E., & Chang, C.C., "Unstable Waves on an Axisymmetric Jet Column," *J. Fluid Mech.*, 65, 541, 1974
- Mchale, R.M., and Nurick, W.H., "Noncircular Orifice Holes and Advanced Fabrication Techniques for Liquid Rocket Injectors Phase I Final Report, NASA CR-108570, Rocketdyne Division of North American Rockwell Corp., Oct. 1970.
- Mchale, R.M., and Nurick, W.H., "Noncircular Orifice Holes and Advanced Fabrication

Techniques for Liquid Rocket Injectors, Comprehensive Program Summary (Phases I, II, III, and IV)," NASA-CR-134315, Rocketdyne Division of North American Rockwell Corp., March 1974.

Nurick. W.H., and Clapp, S.D., "An Experimental Technique for Measurement of Injector Spray Mixing," *AIAA J. Spacecraft*, Vol. 6, No. 11, 1969.

Nurick. W.H., "Orifice Cavitation and Its Effect on Spray Mixing," *ASME J. Fluids Engineering*, Dec. 1976, pp. 681-687.

Nurick. W.H., "Dropmix-A PC Based Program for Rocket Engine Injector Design," JAN-NAF Combustion Meeting, CPIA Publ. 554, 1991.

Roshko. A., "Structure of Turbulent Shear Flows: A New Look," *A.I.A.A. J.*, 14, 1349, 1976.

Rupe. J.H., "The Liquid-Phase Mixing of a Pair of Impinging Streams," Progress Report No. 20-195, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 6, 1953.

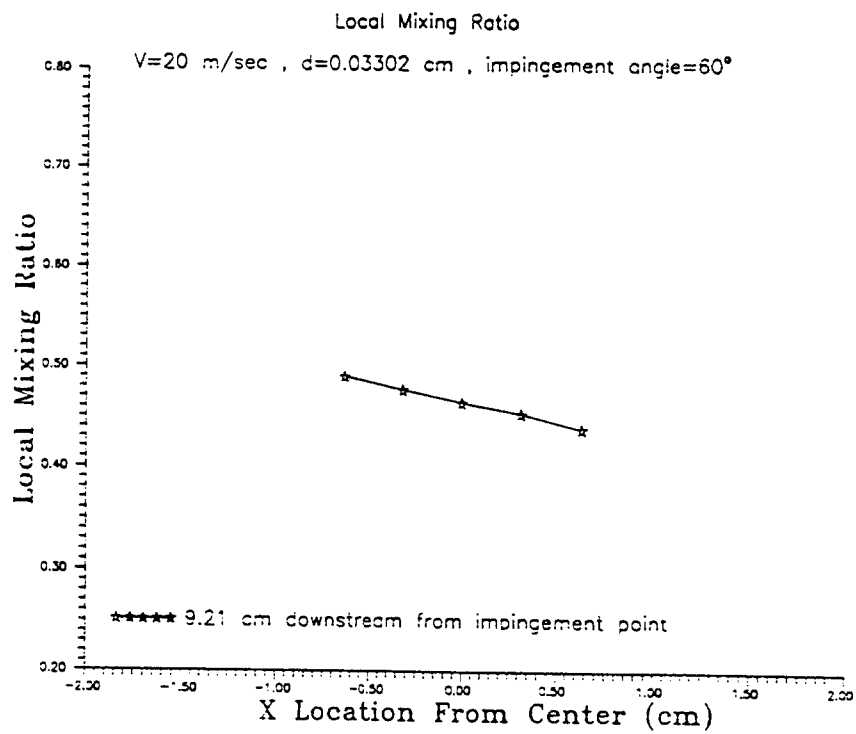
Rupe. J.H., "A Corelation Between the Dynamic Properties of a Pair of Impinging Streams and the Uniformity of Mixture-Ratio Distribution in the Resulting Spray," Progress Report No. 20-209, Jet Propulsion Laboratory, Pasadena, Calif., Mar. 28, 1956.

Ryan. H.M., Anderson. W.E., Pal. S., and Santoro. R.J., "Atomization Characteristics of Impinging Liquid Jets," AIAA paper no. 93-0230, presented at the 31st Aerospace Sciences Meeting & Exhibit, January 11-14, 1993, Reno, NV.

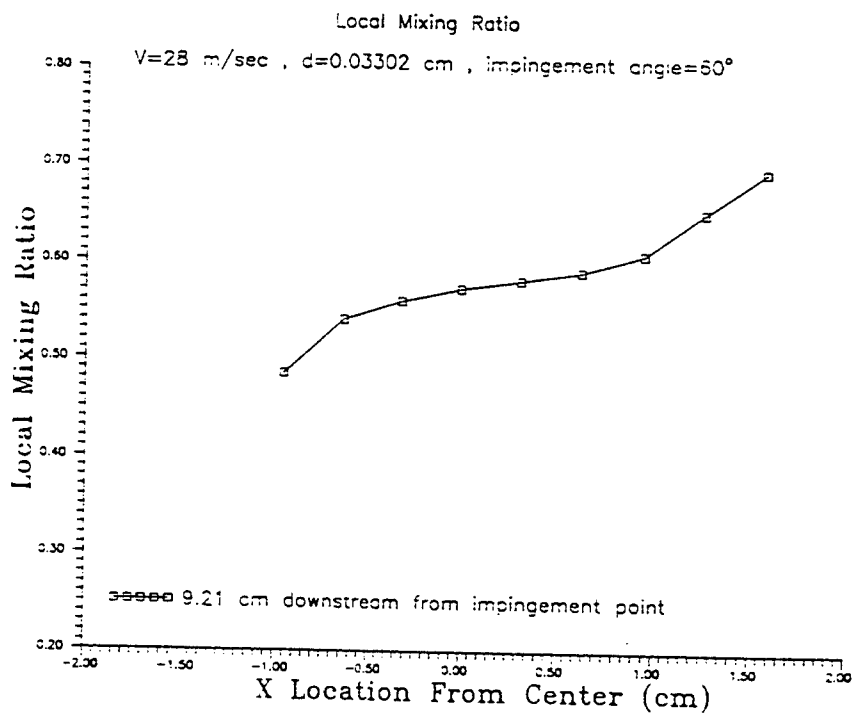
Taylor. G., "Formation of thin flat sheets of water." *Proc. R. Soc. Lond. Ser A.*, 259, 1, 1960.

Vassallo. P., Ashgriz N., and Boorady, F.A., "Effect of Flow Rate on the Spray Characteristics of Impinging Water Jets". *AIAA Journal of Propulsion and Power*, Vol. 8, No. 5, Sept-Oct. 1992.

Yang, H.Q., "Asymmetric Instability of a Liquid Jet." *Phys. Fluids A* 4 (4), 1992.

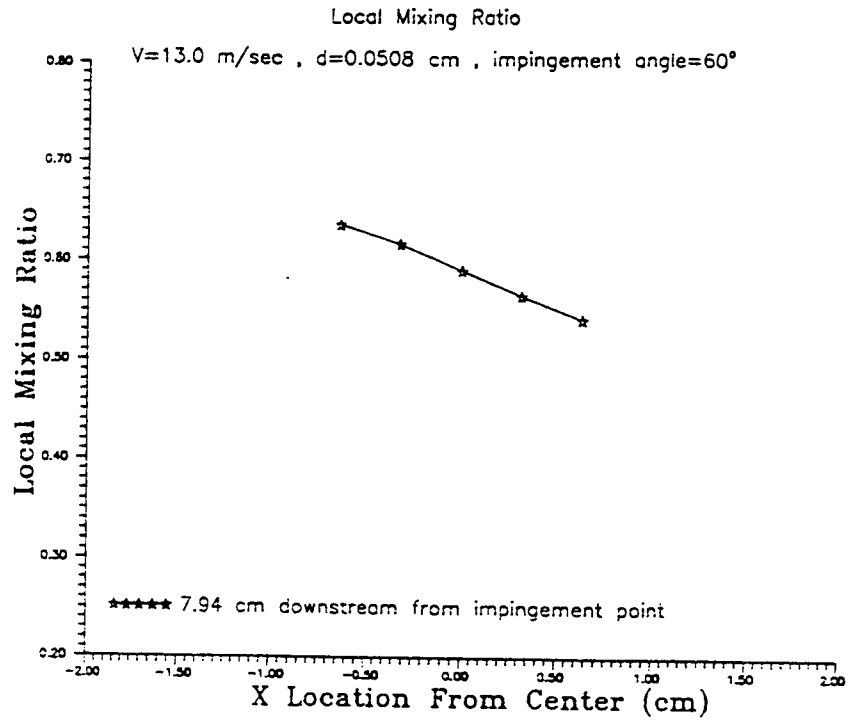


a

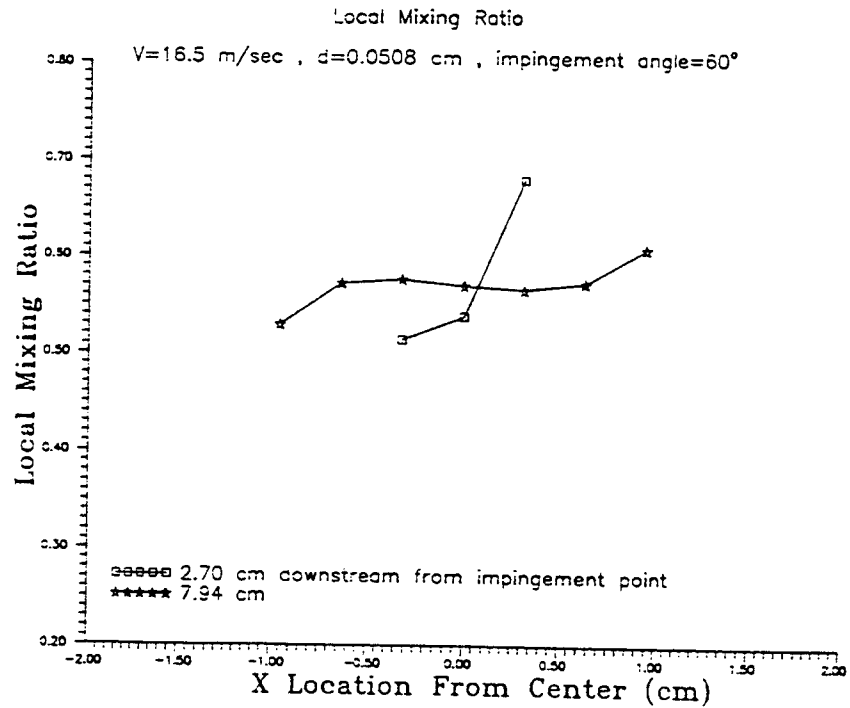


b

Figure 1



a



b

Figure 2

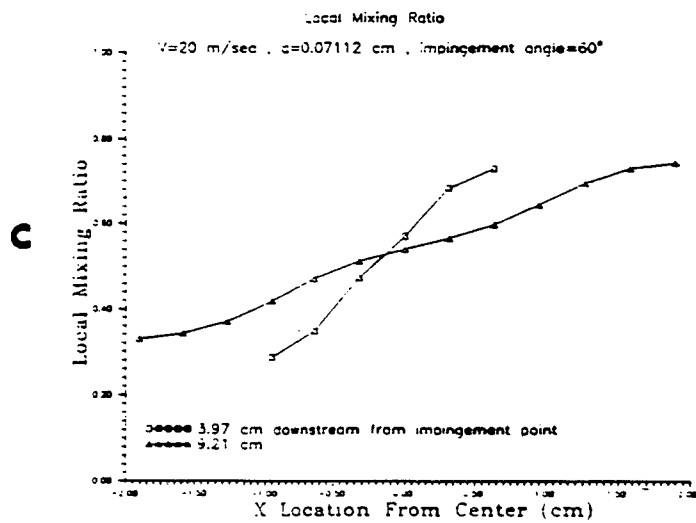
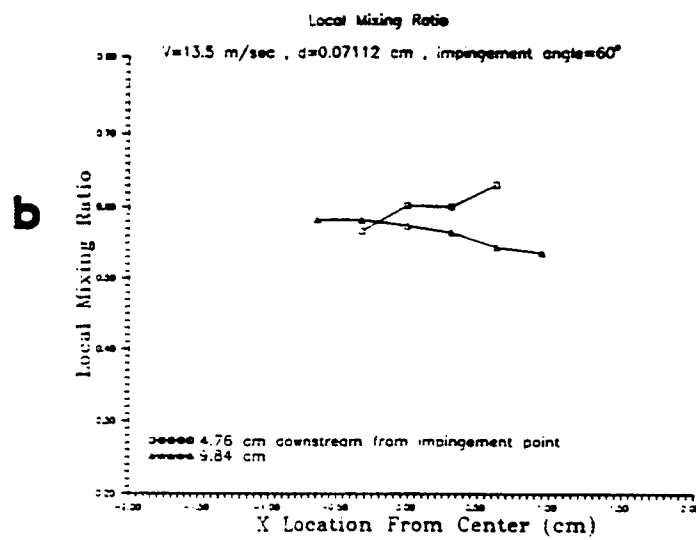
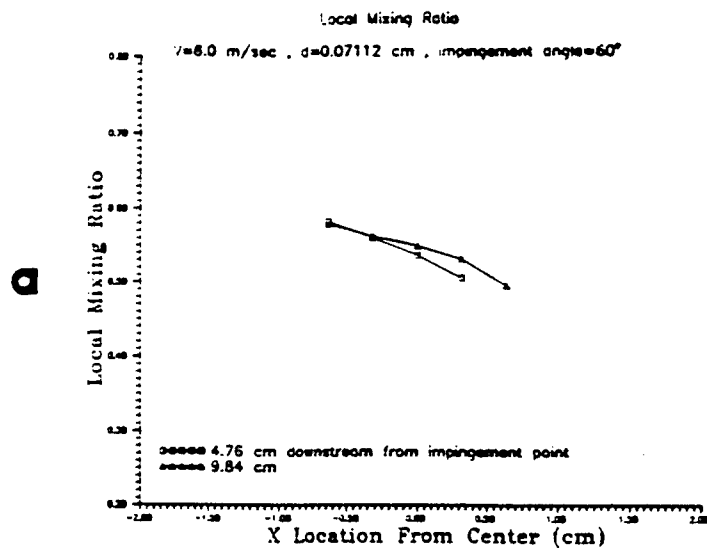


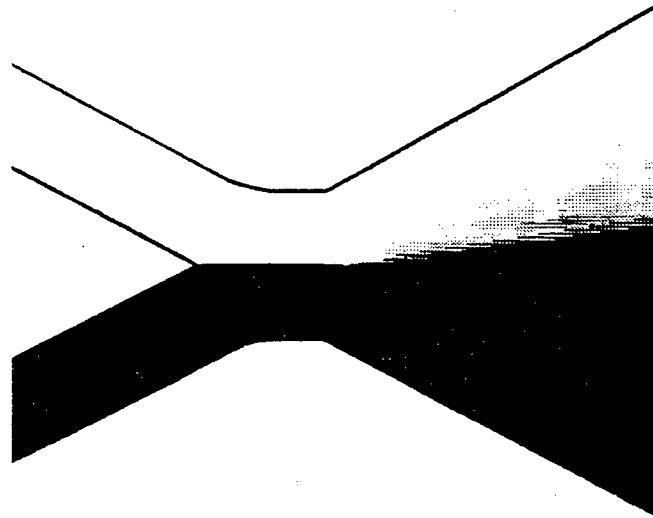
Figure 3

Atomization Modes

Reflective

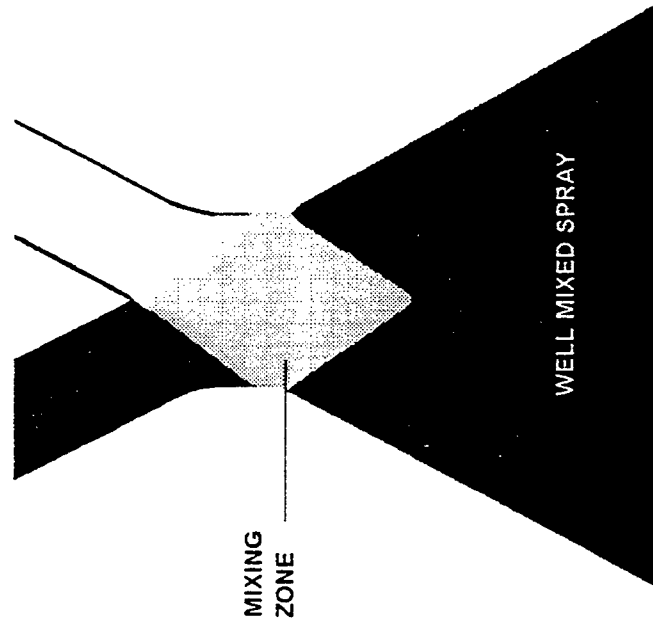
Well Mixed

Transmissive



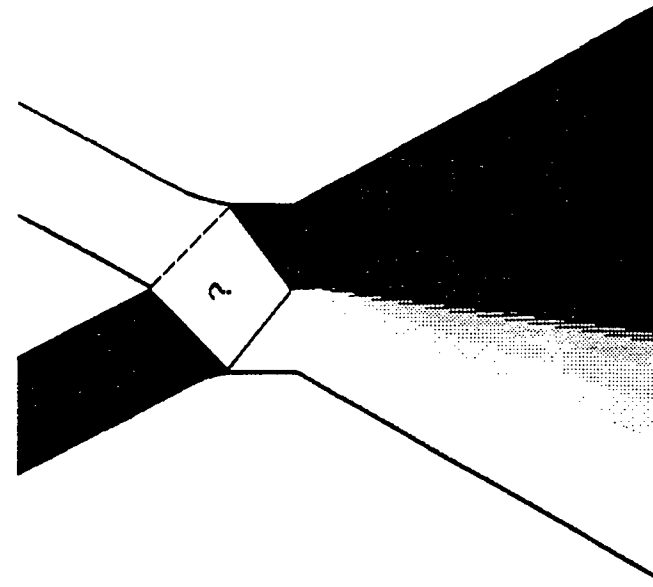
Turbulent Dispersion

(a)



WELL MIXED SPRAY

(b)

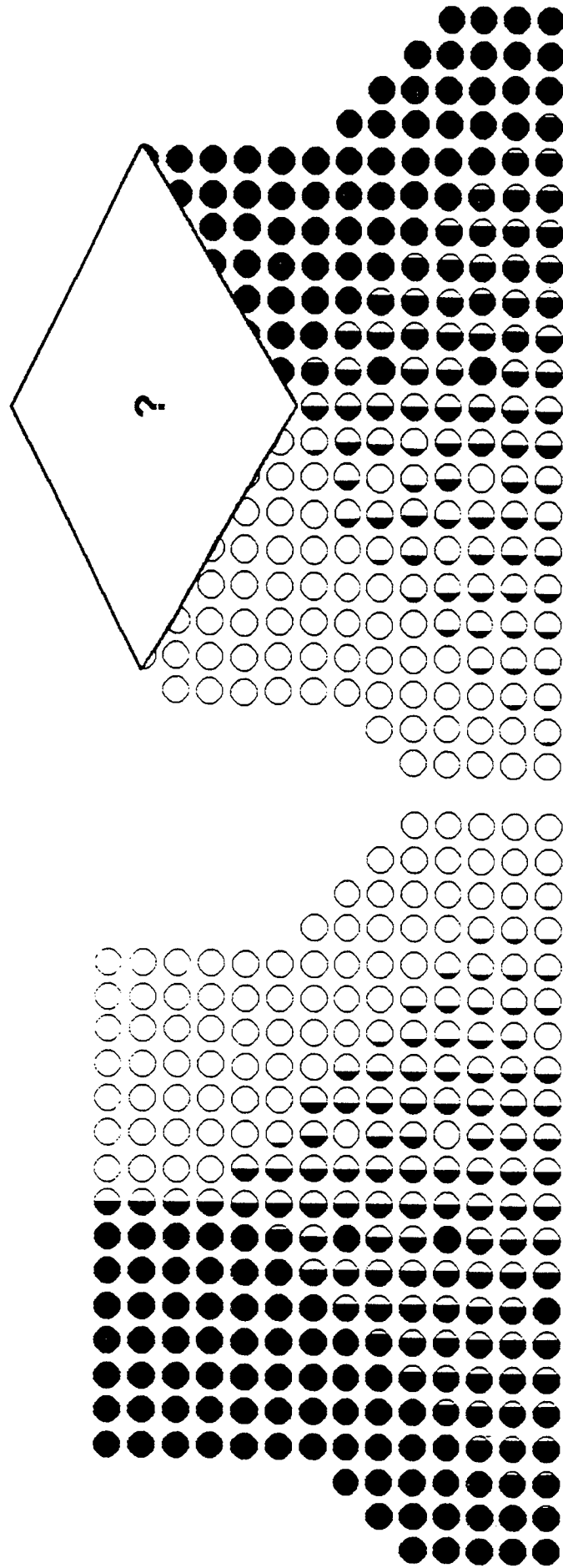


Turbulent Dispersion

(c)

Figure 4

Mixing in Post Atomization Regions



Reflective Impingement

(d)

Transmitting Impingement

(e)

Figure 4 (Continued)

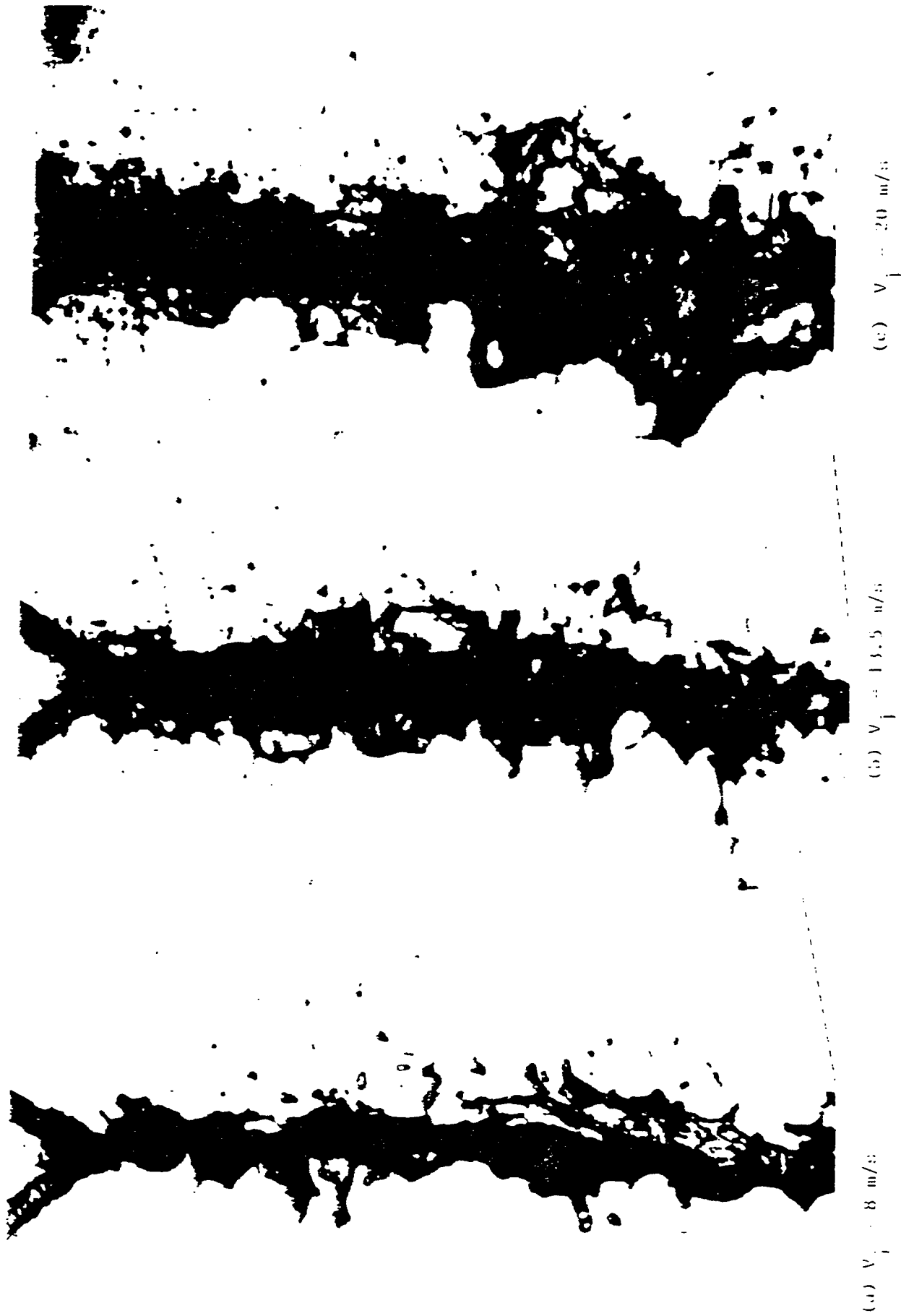
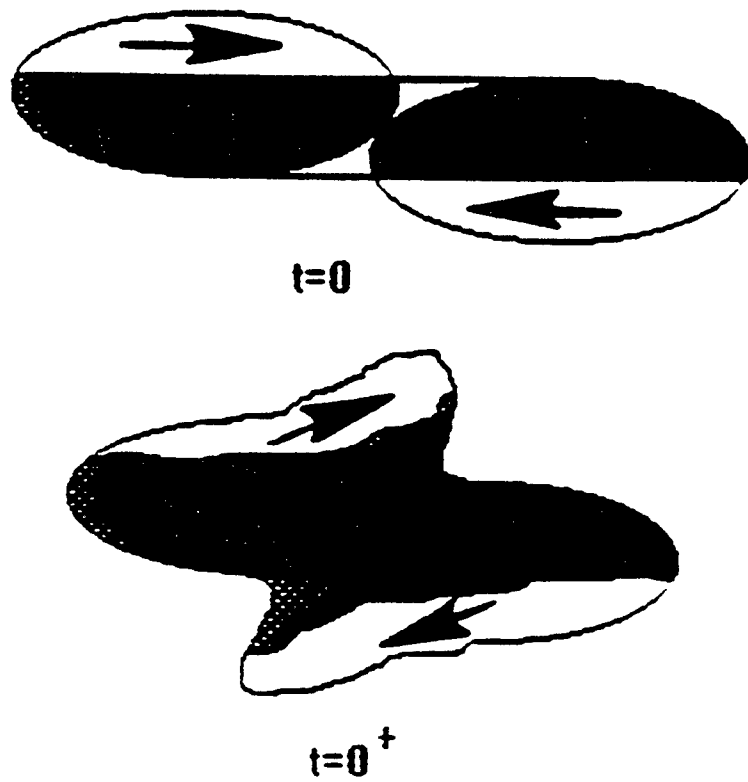


Figure 5: Atomization of an impinging doublet with $D_j = 711 \text{ }\mu\text{m}$, $\alpha = 60^\circ$ and for three jet velocities.

OFF HEAD-ON IMPINGEMENT



**Cross-sectional view of the jets
shortly after impingement.**

Figure 6

VARIATION OF THE EXTENT OF MIXING WITH THE JET VELOCITY

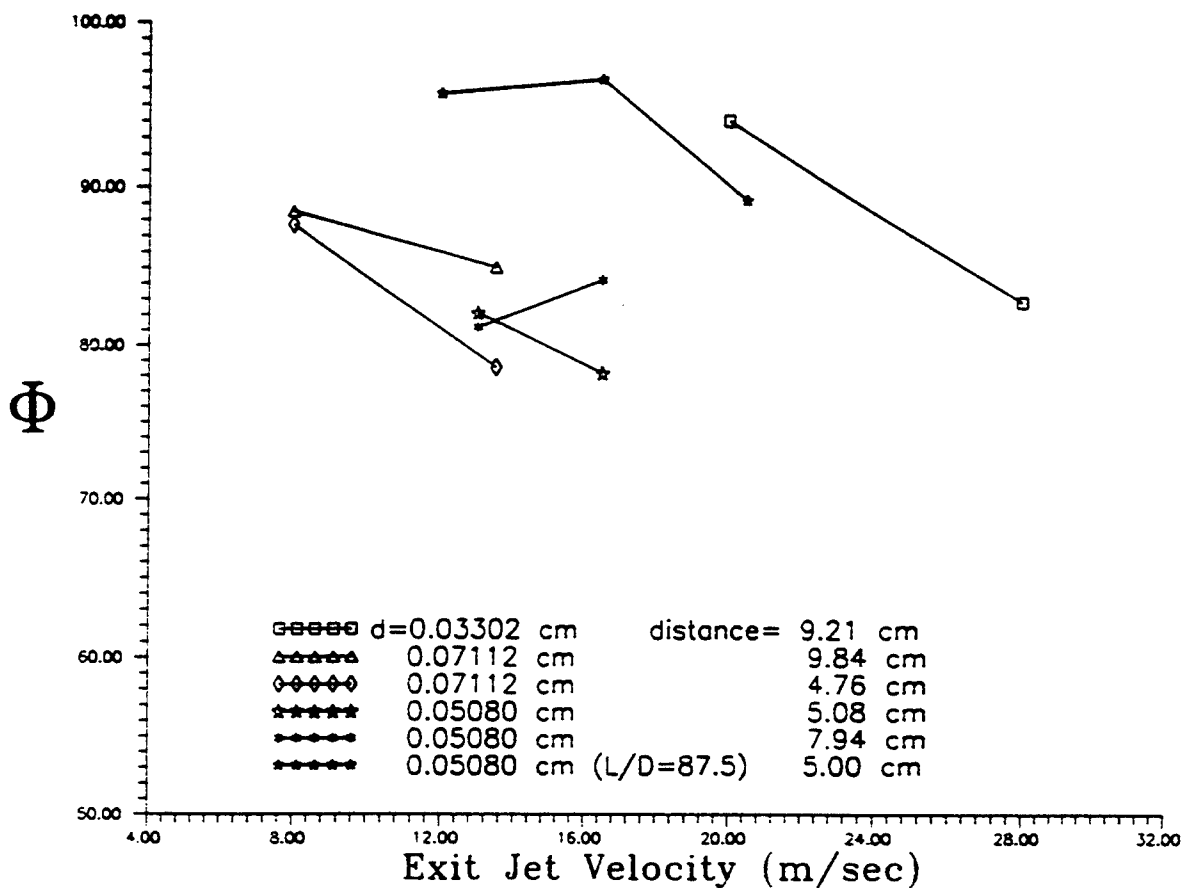


Figure 8

VARATION OF THE EXTENT OF MIXING DOWNSTREAM OF THE IMPINGEMENT POINT

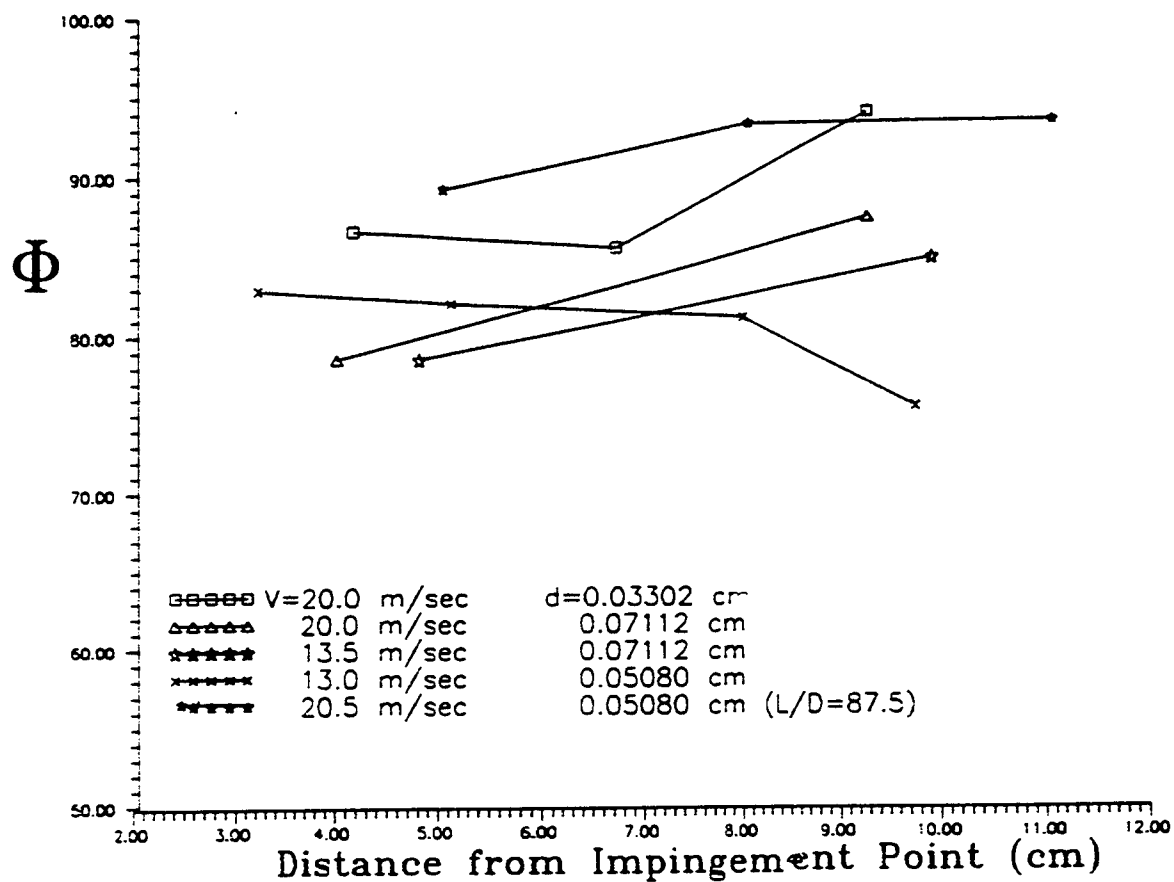


Figure 9

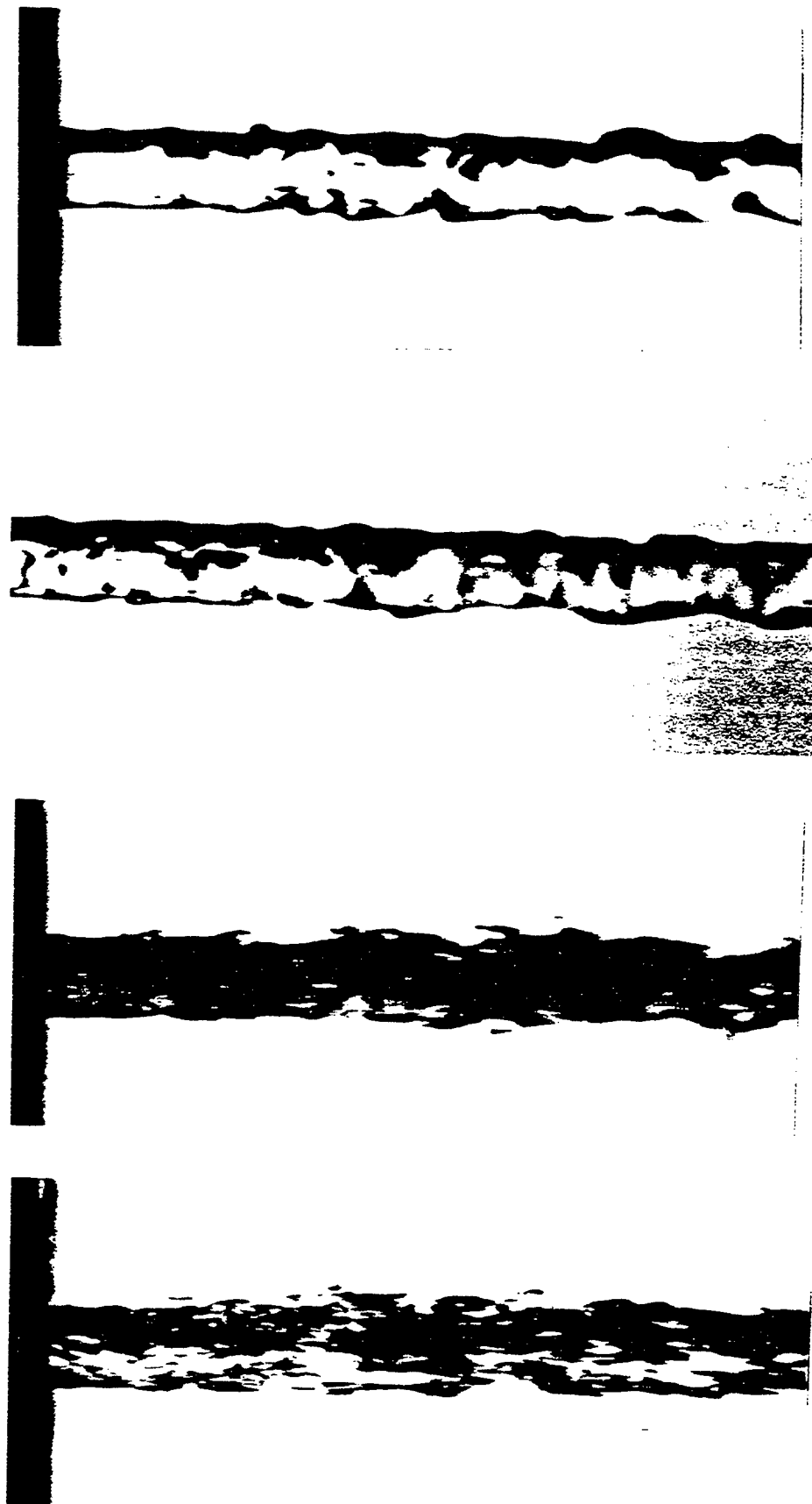


Figure 10

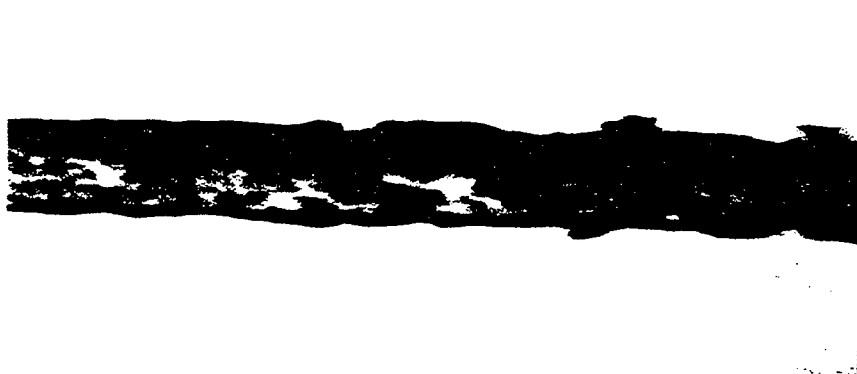
Re = 5000
19000

$D_j = ,6724$

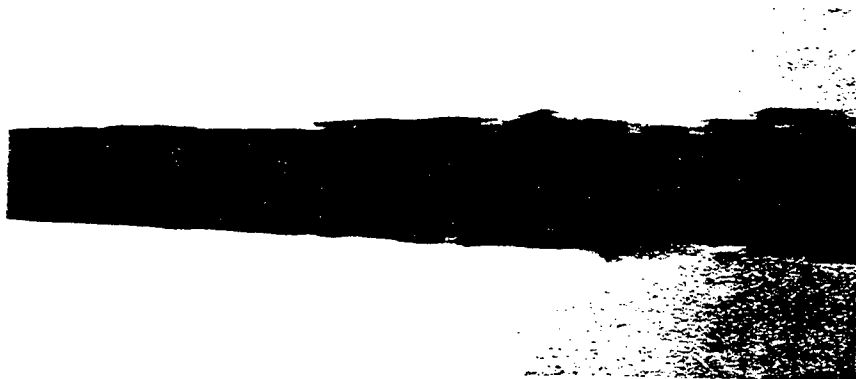
L/D = 10



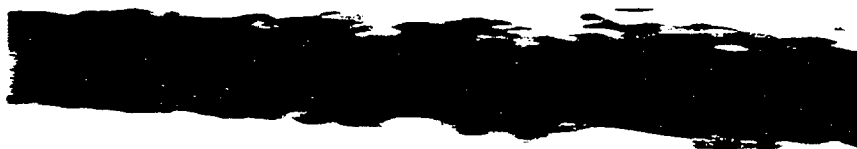
5000



7000



11,000



19,000

Figure 11

$$D_j = .635 \quad L/D = 3$$



Figure 12

Re = 19,000 $D_j = .635$ L/D = 3

AMPLITUDE vs WAVE MODE
SESSION 169 Re= 5000
Dj=0.9899 mm L/D=3

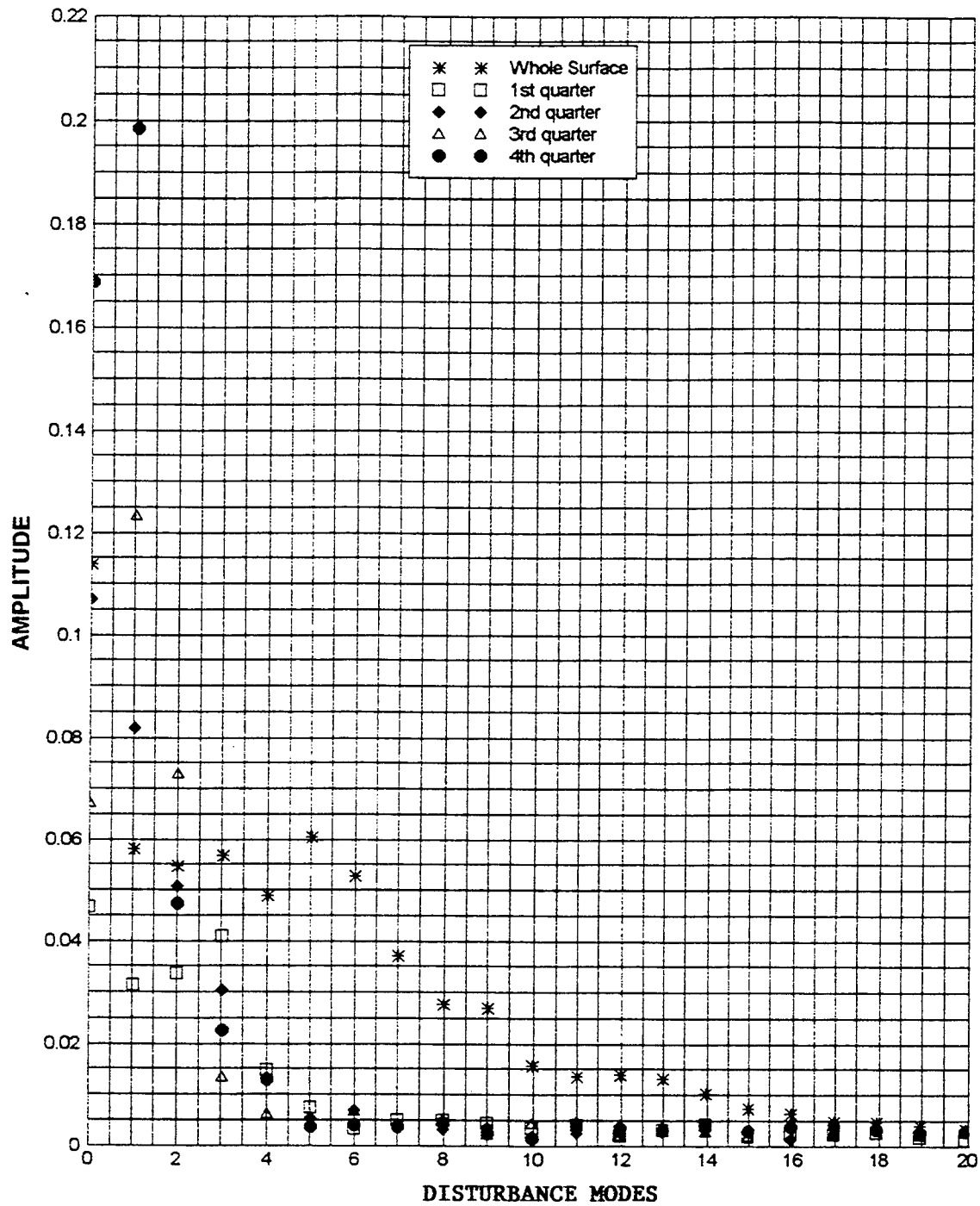


Figure 13

AMPLITUDE vs WAVE MODE
 SESSION 173 Re= 5000
 Dj=.8965 mm L/D=10

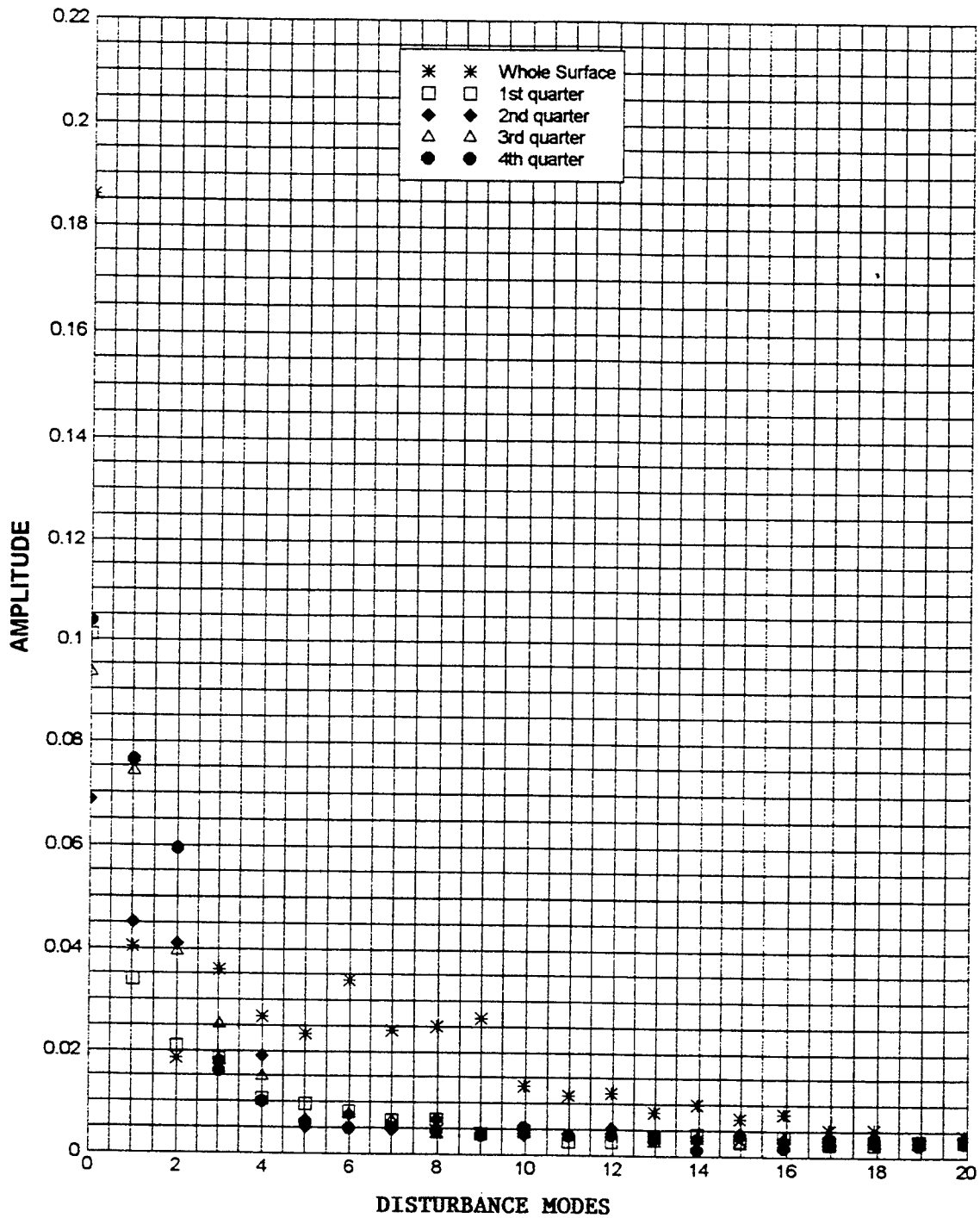


Figure 14

AMPLITUDE vs WAVE MODE
SESSION 174 Re= 18000
Dj=.8965 mm L/D=10

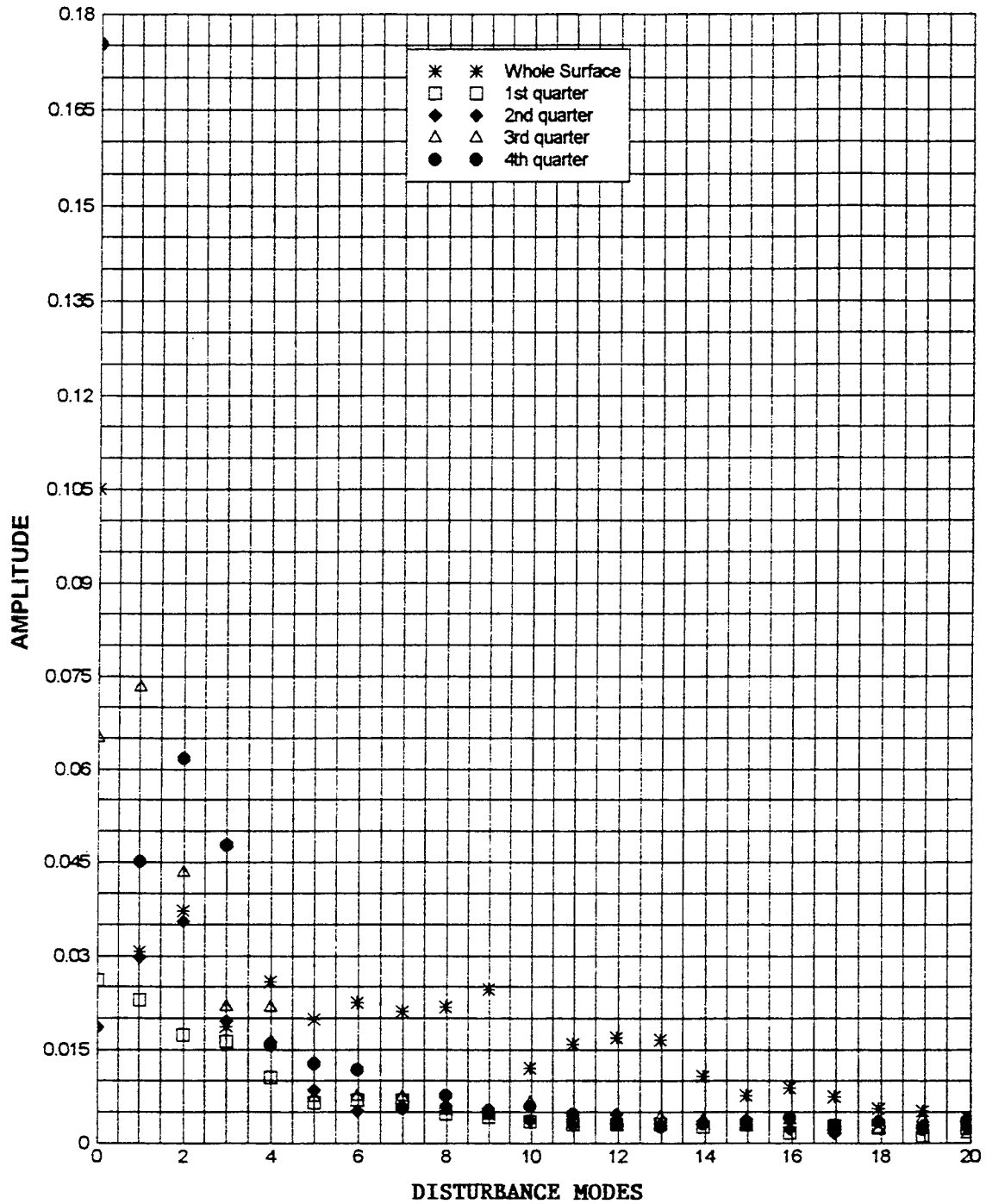


Figure 15

AMPLITUDE vs WAVE MODE
SESSION 176 Re= 7000
Dj=.6350 mm L/D=5

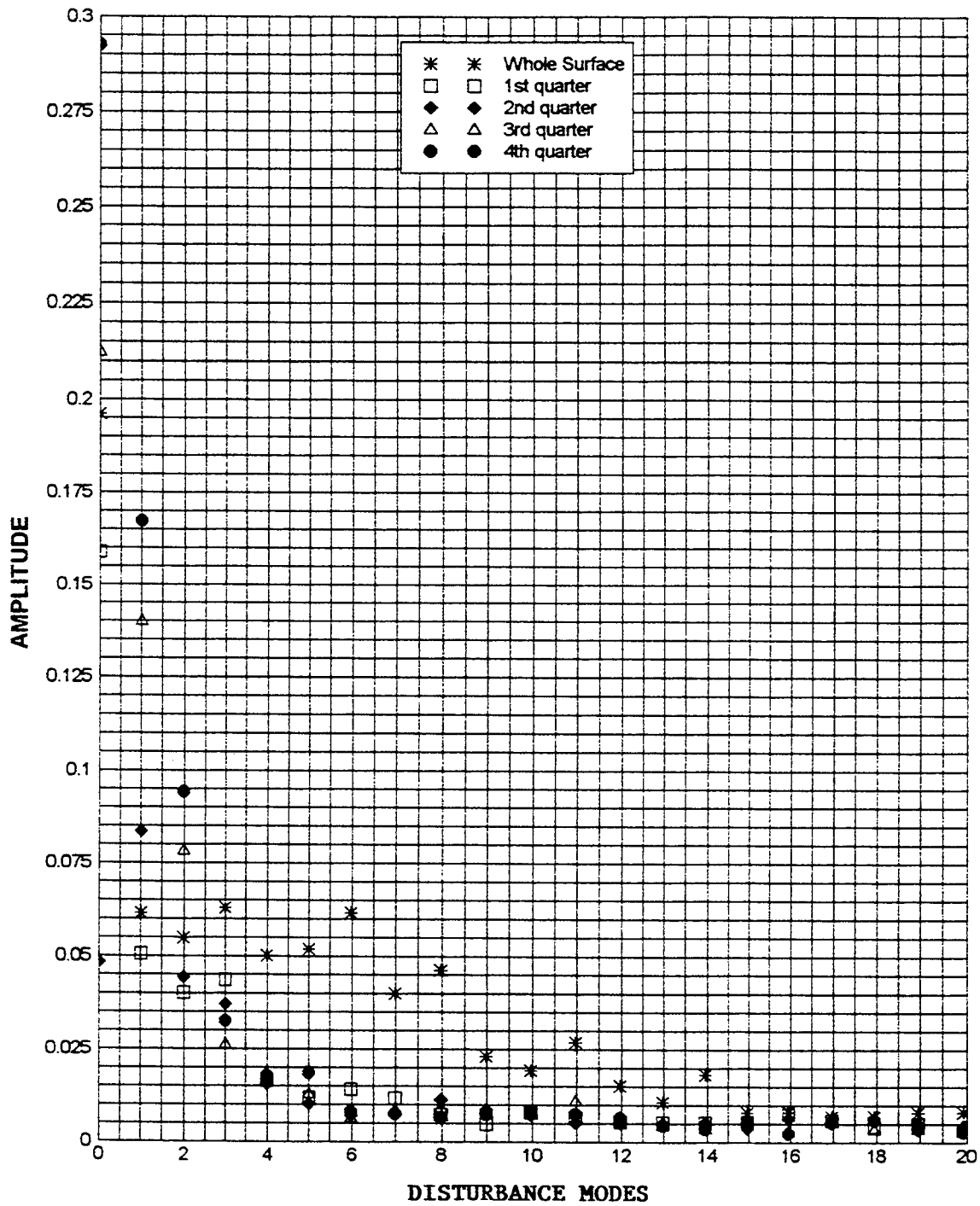


Figure 16

AMPLITUDE vs WAVE MODE
SESSION 197 Re= 18500
Dj=.6350 mm L/D=5

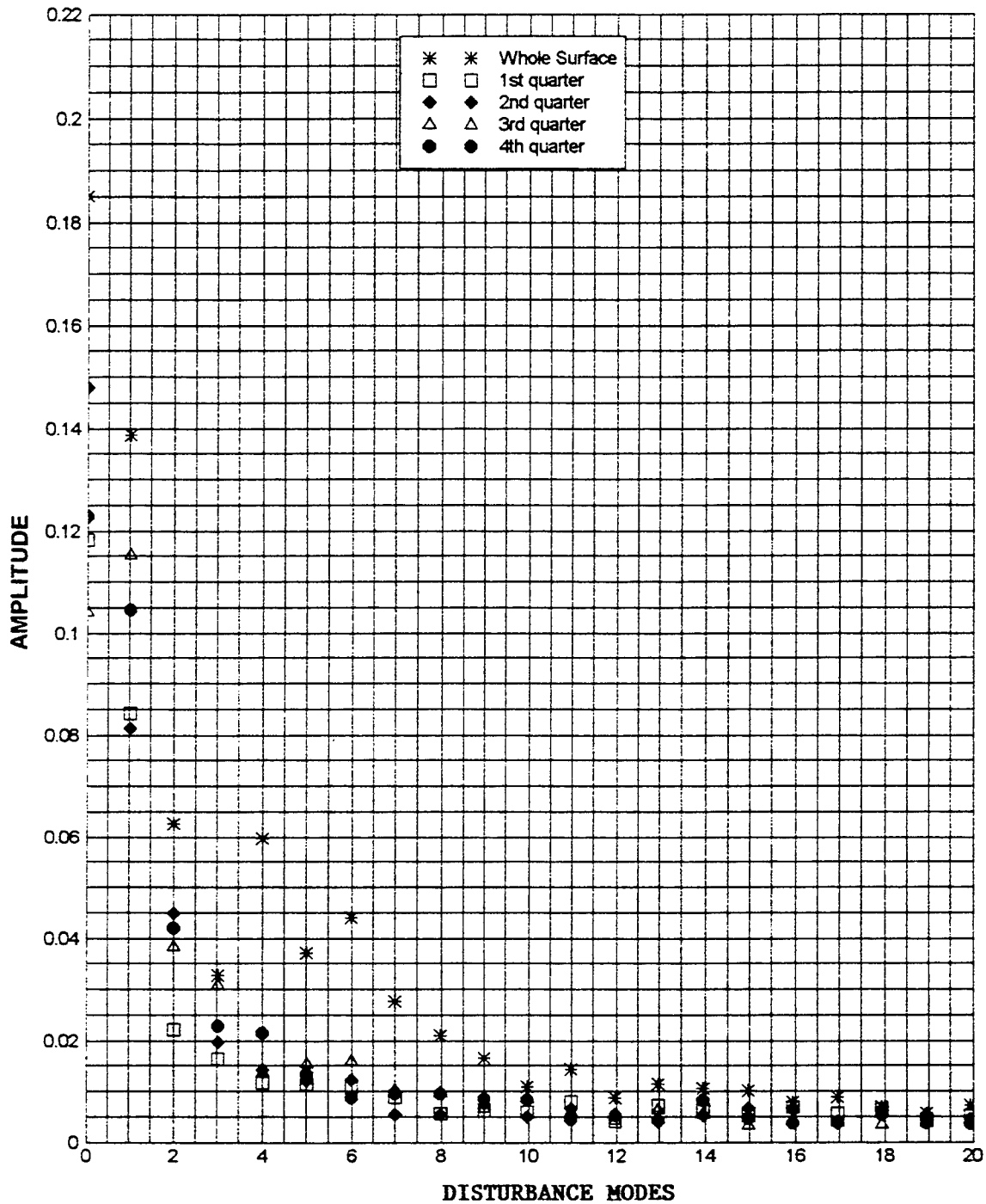


Figure17

AMPLITUDE vs WAVE MODE
 SESSION 203 Re= 7000
 Dj=.6724 mm L/D=10

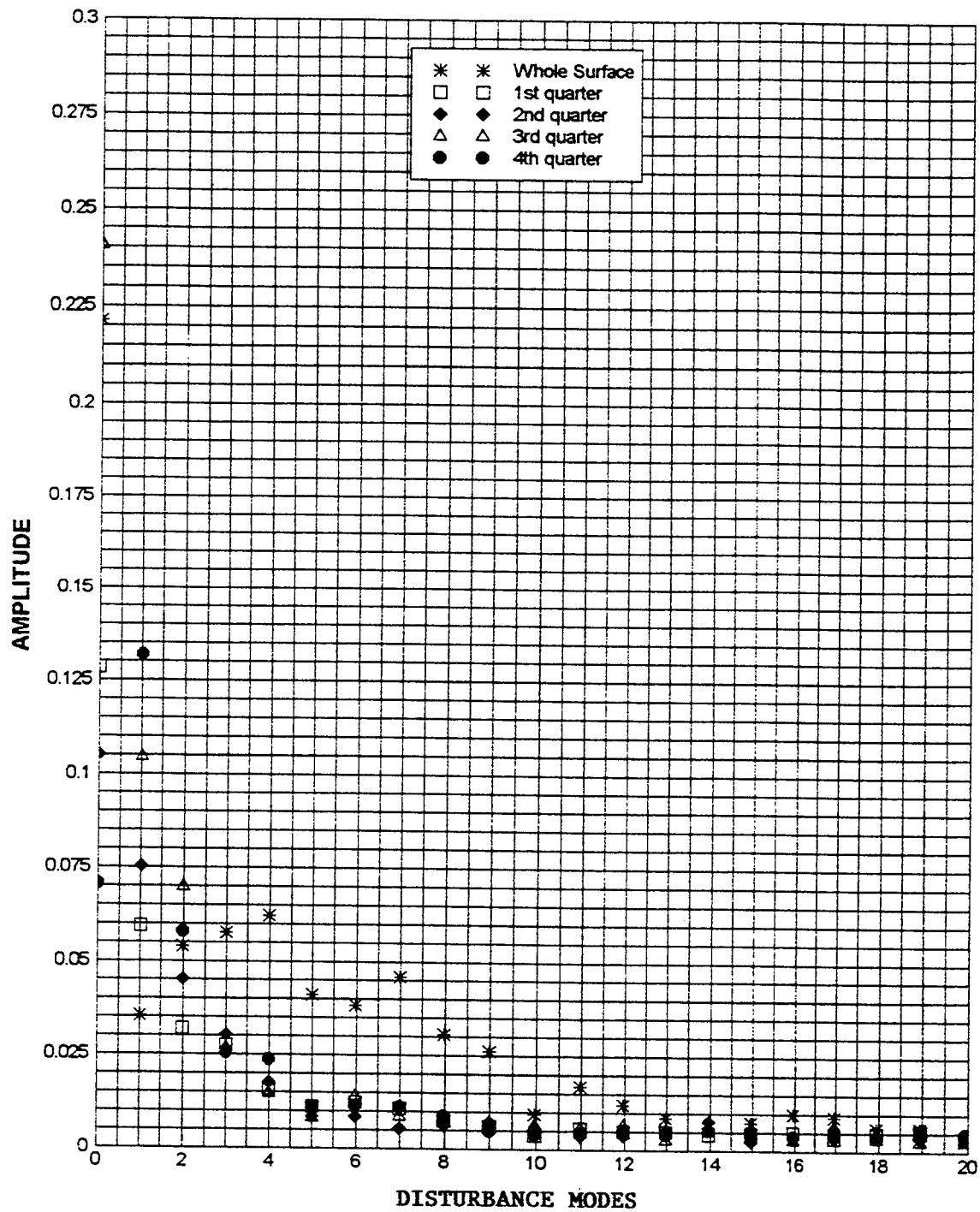


Figure 18

AMPLITUDE vs WAVE MODE
SESSION 209 Re= 19000
Dj=.6724 mm L/D=10

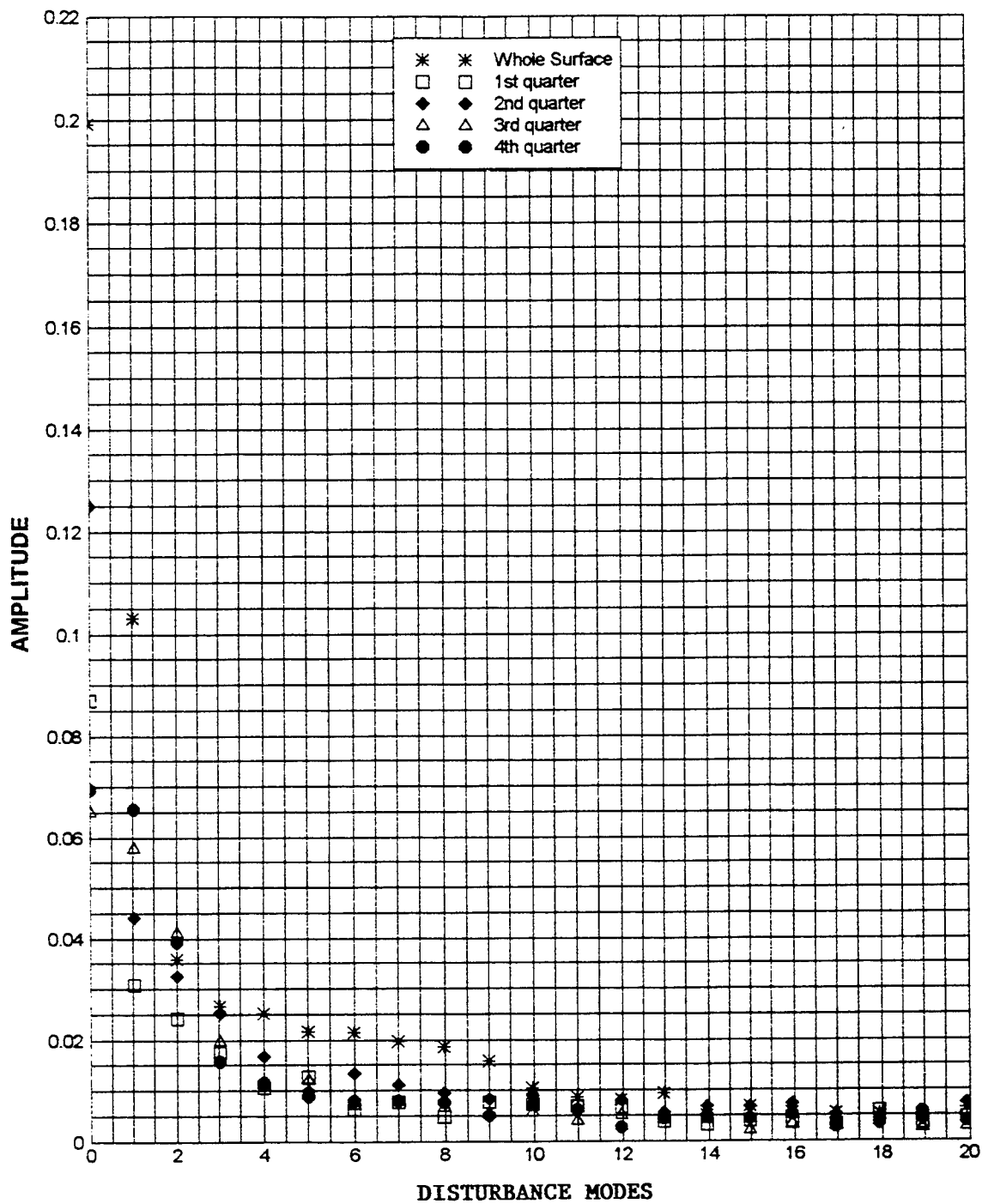


Figure 19

CHARACTERIZATION OF CMOS CIRCUITS USING A HIGHLY CALIBRATED
LOW-ENERGY X-RAY SOURCE

Dr. Raymond D. Bellem
Associate Professor
Department of Electrical Engineering

Embry-Riddle Aeronautical University
Prescott, AZ 86301

Final Report for:
Summer Research Extension Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

Embry-Riddle Aeronautical University

December 1995

CHARACTERIZATION OF CMOS CIRCUITS USING A HIGHLY CALIBRATED LOW-ENERGY X-RAY SOURCE

Raymond D. Bellem
Associate Professor
Department of Electrical Engineering
Embry-Riddle Aeronautical University

Abstract

This paper describes continuing work on a two-phased program to investigate the measurement and calculations of x-ray radiation transport in silicon: and study the effects of ionizing x-ray radiation on Complementary Metal Oxide Semiconductor (CMOS) devices and circuits fabricated using standard commercial processing steps.

The attenuation of a 20-160 keV X-ray spectrum through a silicon target was calculated using the CEPXS/ONELD transport code. Both the transmitted and incident spectra were measured using a HPGe spectrometer. The calculated and measured data were compared and show excellent agreement - leading to the conclusion that the dose measurements and the transport calculations were well within the accuracies required by ASTM F-1467 and Mil-Std 883 to accomplish accurate ionizing radiation research and testing on microelectronic and photonics devices and circuits.

A test chip was designed using a hardened circuit design methodology which yield radiation tolerant IC's when using a commercial CMOS fabrication process. Several CMOS test circuits and P and N-channel MOS devices were irradiated up to dose levels of 300krad(SiO₂) at a dose rate of 100 rad/sec in accordance with MIL-STD-883D, Test Method 1019.4 and ASTM F-1467. The Phillips Lab X-ray source calibrated as part of this effort was used to provide an ionizing radiation environment. Radiation test data is presented on the three circuits - a 21 gate delay chain, an XROM sense amplifier and a 4-bit shift register. All test circuits operate effectively through 100krads and show less than 20% degradation after 300krads. Reentrant (annular) structures show little threshold shift or leakage current after 300krads. Edge transistors show little threshold shift at 100krads with increasing shifts up to 0.3 volts at 300krads with leakage currents increasing several orders. Field oxide devices show significant leakage current after 100krads.

CHARACTERIZATION OF CMOS CIRCUITS USING A HIGHLY CALIBRATED LOW-ENERGY X-RAY SOURCE

R.D. Bellem

INTRODUCTION

The use of commercial technologies in military systems has received a great deal of attention and evaluation by the DoD over the past several years. Budgetary constraints and technological requirements brought on by the changing threat and mission requirements of the military have led to a reevaluation of investments in mil spec equipments for many applications. This has led to R&D efforts within DoD to adapt semi-tolerant commercial and dual-use technologies wherever possible in military electronic systems. This approach will not only have tremendous cost advantages but will also draw upon the strengths of rapid advances in commercial electronic technologies, especially those experienced in integrated circuits.

The radiation hardening requirements on both space and missile systems has changed due to recent changes in the threat scenario and the strategic posture of the United States. Radiation tolerant integrated circuits however, continue to be a challenge for certain well-defined hostile environments such as the natural radiation environments of space. With advances in IC technologies has come an increased sensitivity to the space environment. This growing sensitivity of electronics to natural radiation is due to device and density scaling inherent in miniaturization, as well as radiation introduced during state-of-the-art integrated circuit processing steps. These factors are also becoming a major concern for commercial vendors of IC's as well. Therefore, radiation effects on electronics research has shifted from the arena of high total dose and high dose rates to low dose rates, moderate total dose and single event effects.

CMOS technology finds widespread application in military systems. Most signal and data processing functions along with memory are implemented using CMOS. The use of commercial CMOS circuits for application in military systems such as satellites is of great interest. The research being done in this effort looks at the effects of ionizing radiation on devices and circuits fabricated with a standard commercial CMOS process. The goal of this research is to determine the radiation tolerance of integrated circuits designed using a circuit design methodology while using standard IC CMOS fabrication techniques.

Part I

Analysis of X-Ray Transport in Silicon For Energies In The 20-160 keV Range

The calibration work on the Low Energy X-Ray (LEXR) facility at the Phillips Laboratory has been ongoing for approximately three years now. Much of the prior research work has been reported in the literature [1-4]. The LEXR provides a well characterized radiation environment for evaluating the capabilities of radiation transport codes in the 20-160 keV energy range. This report addresses the work done over the past year looking at the accuracy of the electron-photon transport codes PHOTCOEF and CEPXS/ONELD and refining the spectrometer measurement techniques developed earlier [2]. PHOTCOEF and CEPXS/ONELD are commercially available electron-photon transport codes used for shielding and depth-dose calculations in any material structure[5]. Since most of the research and testing will be done on silicon structures, an experiment was developed to use the attenuation of a measured X-ray spectrum thru a silicon absorber as a means to measure the accuracy of code calculations. The spectral attenuation caused by a silicon target, was calculated using PHOTCOEF and CEPXS/ONELD using the measured incident spectrum as the input spectrum required by these codes. Both the transmitted and incident spectra were measured using a high purity germanium spectrometer system [3]. The measured incident spectra is used as the input spectra for PHOTCOEF and CEPXS/ONELD for the transmission calculations.

APPROACH

The approach consists of the following steps:

- 1). Measure the baseline dark current (leakage) of the detector (tube current off).
- 2). Measure the backscatter signal from the rear wall of the LEXR facility.
- 3). Measure the incident spectrum of the x-ray tube at the silicon surface.
- 4). Measure the transmitted spectrum exiting the silicon wafer.
- 5). Use the spectrum measured in 3) as the input spectrum for the CEPXS and PHOTCOEF calculations.
- 6). Calculate the exiting spectrum using CEPXS and PHOTCOEF.
- 7). Analyze and compare the data from 4) and 6).
- 8). Assess the accuracy of CEPXS and PHOTCOEF for transport and depth-dose calculations in silicon material systems.

EXPERIMENTAL SET-UP

The LEXR Facility shown in Fig. 1 was used for all experimental work. Details of the facility have been discussed in previous publications [2,3].

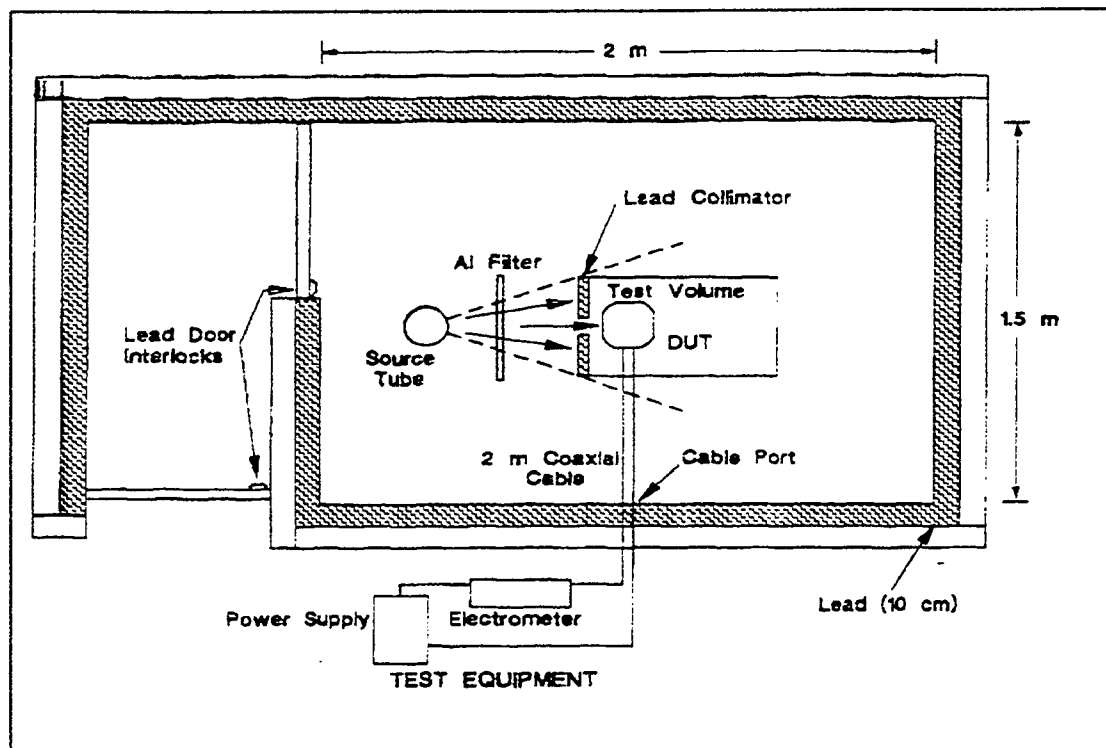


Figure 1. Schematic of Low Energy X-Ray Facility

Figure 2 shows a cut away of the x-ray tube, the collimated path of the x-ray beam, thru the silicon wafer, followed by the cryogenic chamber containing a high purity germanium detector. The germanium detector is the photon detector of a high resolution EG&G ORTEC x-ray spectrometer. Details of the spectrometer measurement techniques were discussed in [4]. Collimator 1 is a 5.1 cm Pb slab with a 12.7 mm diameter aperture. Collimator 2 is a 1.27 cm Pb slab with an entrance aperture of 1.8 mm diameter and exiting aperture of 0.34 mm diameter. Collimator 3 is a 1.27 cm Pb slab with a 3.17 mm diameter aperture. Collimators 1 and 3 were large aperture aperture structures designed primarily to reduce photon scattering from the lead walls of the chamber, while collimator 2 primarily limited photon flux. A laser beam was used to align the detection system, collimators, and x-ray tube by matching the incident and reflected beams. In addition, low x-ray tube currents were used to keep the measurement systems dead time at or below 15 percent.

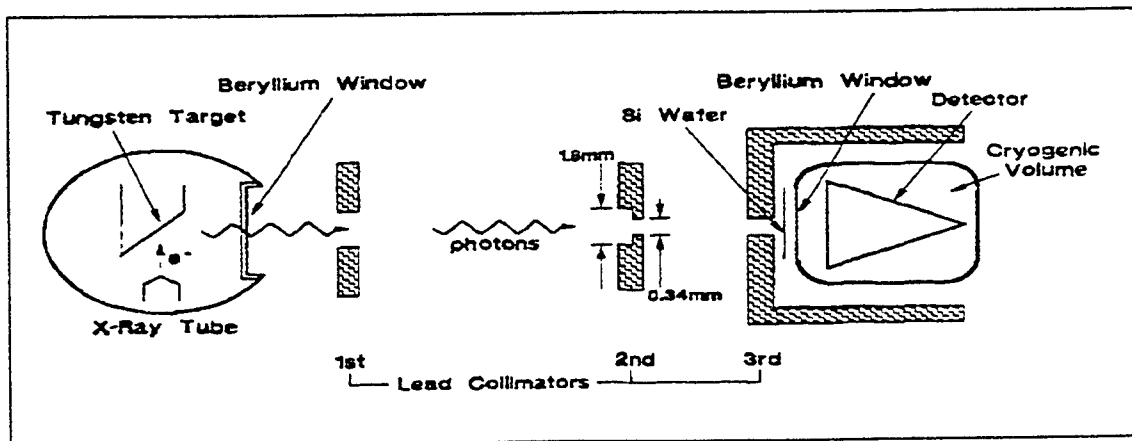


Figure 2. Cut-Away of the photon source, path, collimation, target and detection system.

MEASUREMENTS

The first two steps involves the measurement of the background and back-scatter from the walls of the x-ray chamber. Figure 3 shows the results of this measurement. Photon counts were corrected for different measurement times. Counts of 500 in $\#/\text{keV}$ can be neglected when compared to actual spectral measurements - which are orders of magnitude higher in this energy range. The peak at first energy bin can also be neglected as it is anomalous with the method of data reduction and is not representative of a high photon count in this energy bin.

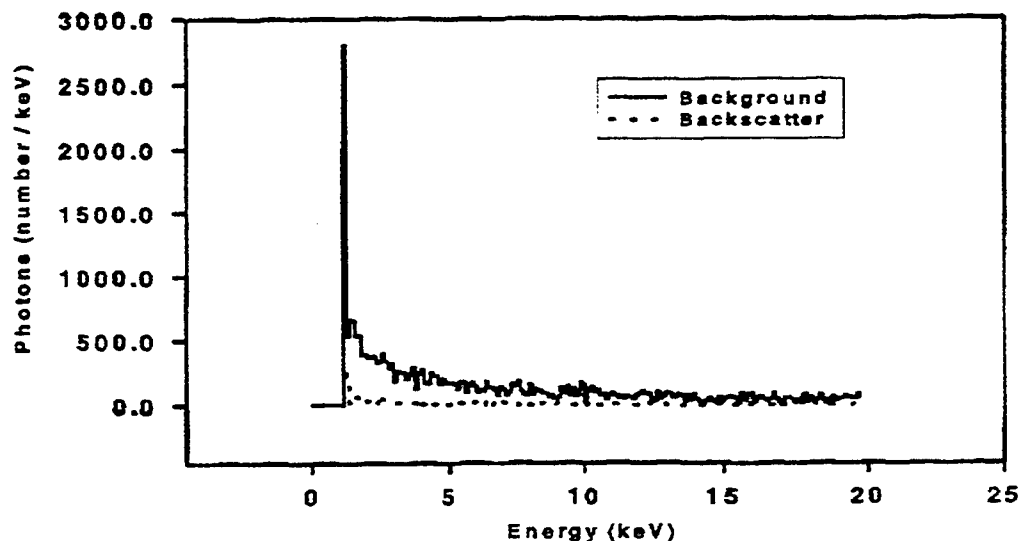


Figure 3. Background Measurements and Back-scatter Measurements From Walls of LEXR.

Step 3 involves measuring the incident LEXR spectrum at the surface of the silicon target. In order to decrease the photon flux, the distance from the x-ray tube to the silicon target was selected at 82cm. The x-ray tube current had to be decrease to values in the microamp range in order to decrease the spectrometer dead time below 15%. Incident spectrum measurements were made at tube voltages of 20,40,50,60,80,100,120,140 and 160 kV. The incident spectra were measured and recorded for use in the transport codes. Figure 4 shows two representative plots of the data gathered. Note that the L and K-lines appear very prominently and are very important for this study.

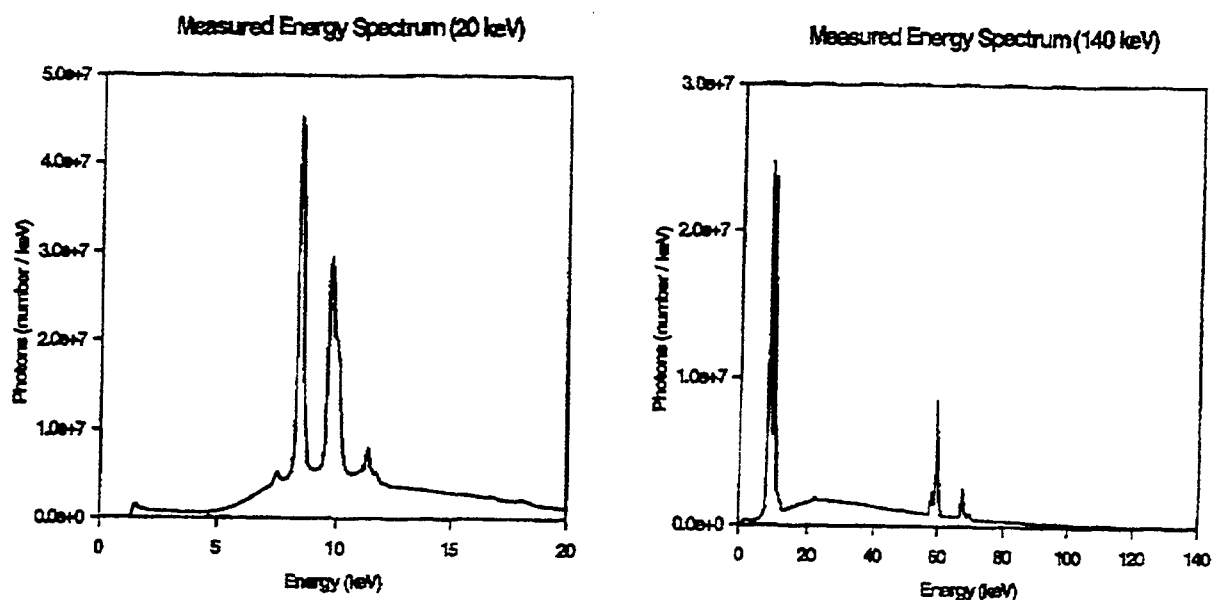


Figure 4 Measured Incident LEXR Spectrum at 82 cm.

The last measurement step involves measuring the attenuated spectrum after transmission through the silicon. An IC wafer grade silicon (110) target with a thickness of 0.0533 cm. was used. Figure 5 shows plots of the measured incident and transmitted energy spectrum for the 20 and 140 kV spectra. Again, transmission data was taken for all spectra from 20 kV to 160 kV.

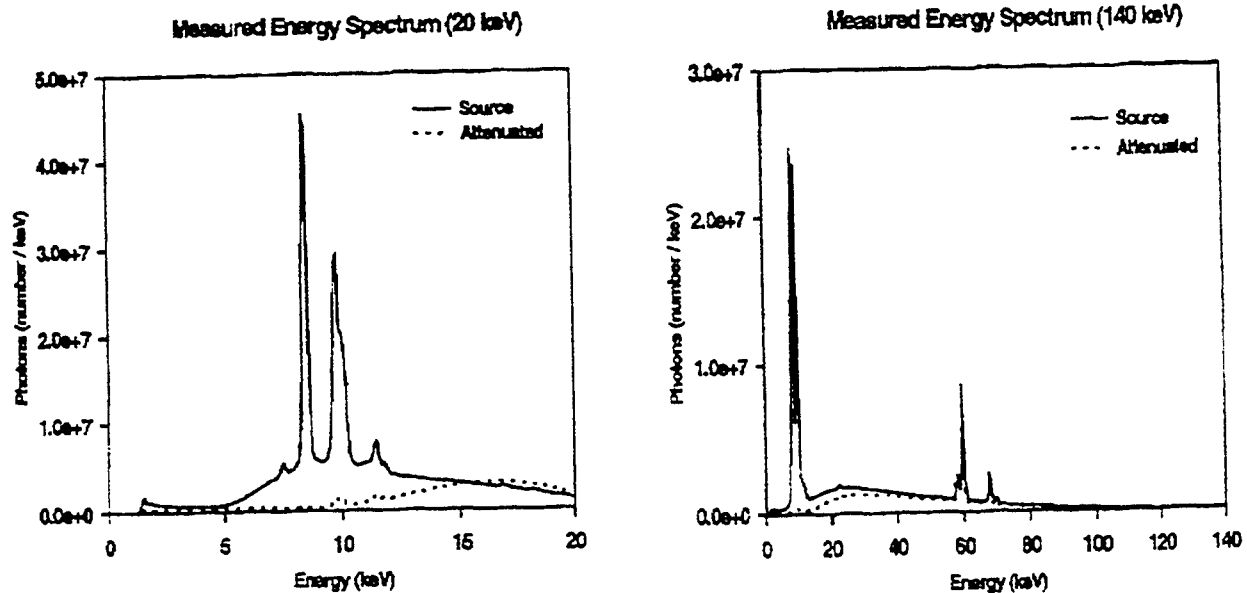


Figure 5. Plot of Transmitted spectra overlaid on the incident spectra plots for 20 and 140 kV.

CALCULATIONS

A description of Photcoef and CEPXS can be found in [2]. Figure 6 depicts the steps involved in making the transport calculations. Both Photcoef and CEPXS requires a normalized spectrum for use in all calculations. Figure 6 shows the steps accomplished in accomplishing all calculations.

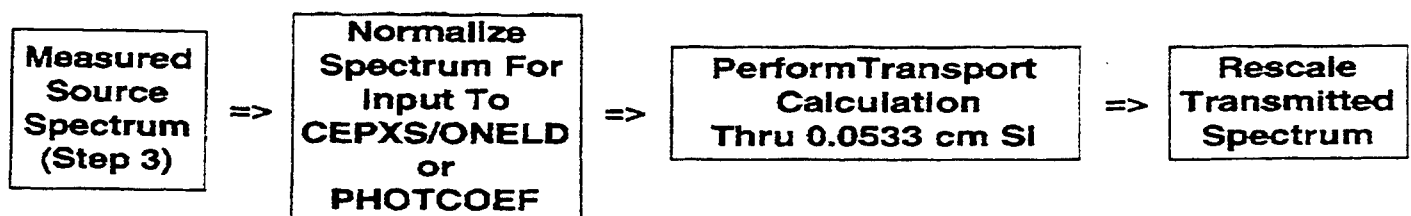


Figure 6. Transport calculation technique.

DATA COMPARISON AND ANALYSIS

The next step in this study is to compare the results of step 4 and steps 5 & 6. Figure 7 shows example plots of the measured and CEPXS/ONELD calculated transmitted spectra for 20 kV and 140 kV. Figure 8 shows a similar plot for the Photcoef calculated spectra.

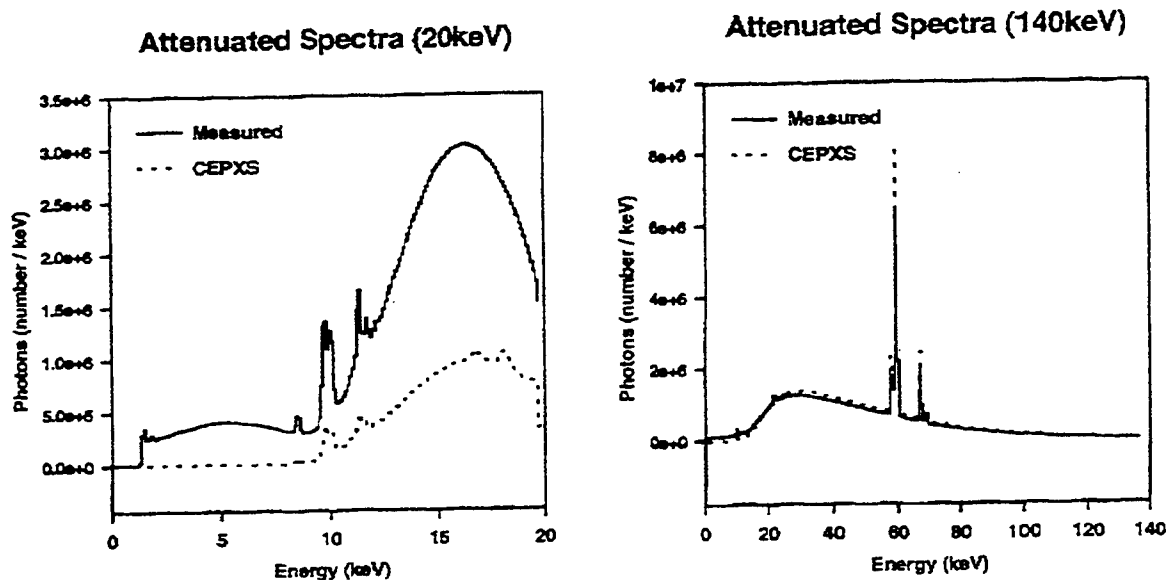


Figure 7. Comparison of measured and CEPXS/ONELD calculations.

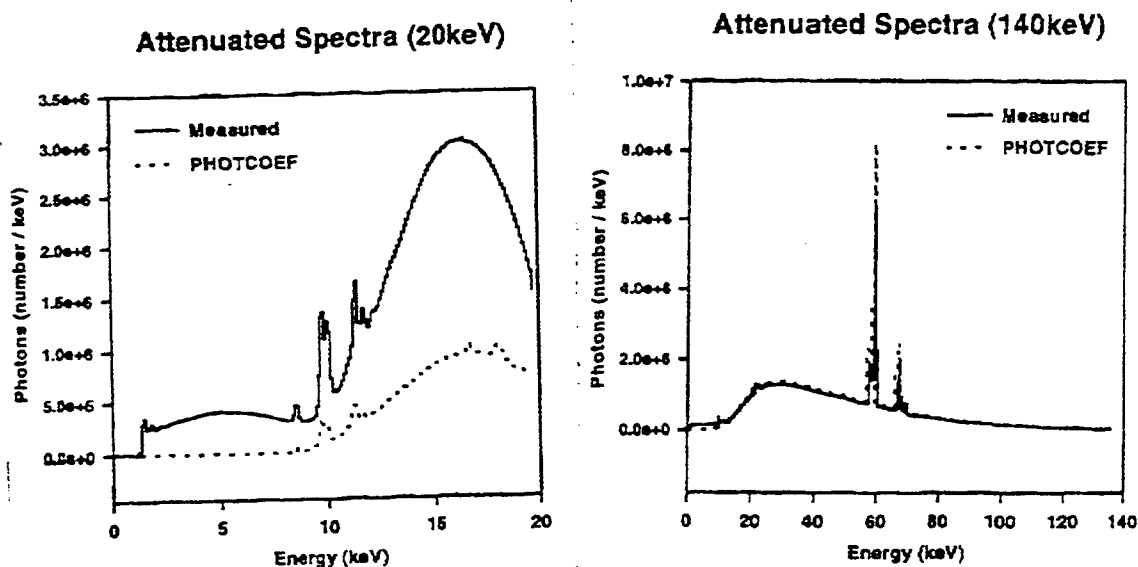


Figure 8. Comparison of measured and Photcoef calculations.

SUMMARY OF COMPARISON DATA

The last step will be a comparison of the data over the full range of the X-ray source. The comparisons will be over the total spectrum and then the continuum, L and K line data will be broken out for further comparison and analysis. Figure 9 shows a comparison of percent error resulting from comparing both CEPXS and Photcoef calculations with the measured transmitted data across the full spectrum.

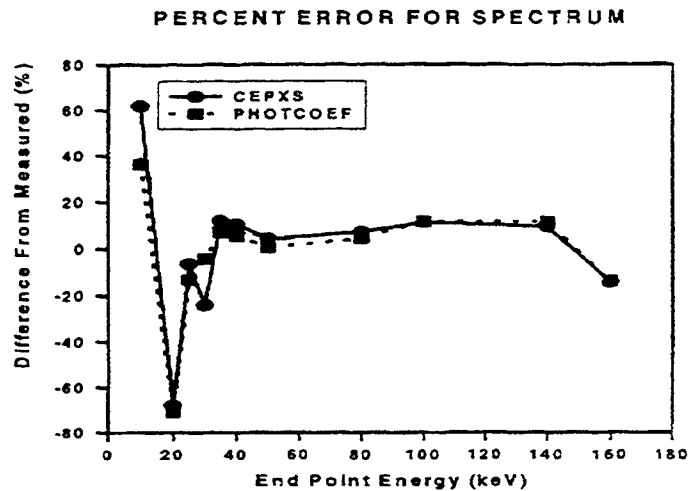


Figure 9. Error comparison for calculations over full spectrum.

In order to better understand the introduction of errors in the calculations or the measurements, it is instructive to separate the continuum and the line spectra. Fig. 10 shows the error of the two codes for the continuum.

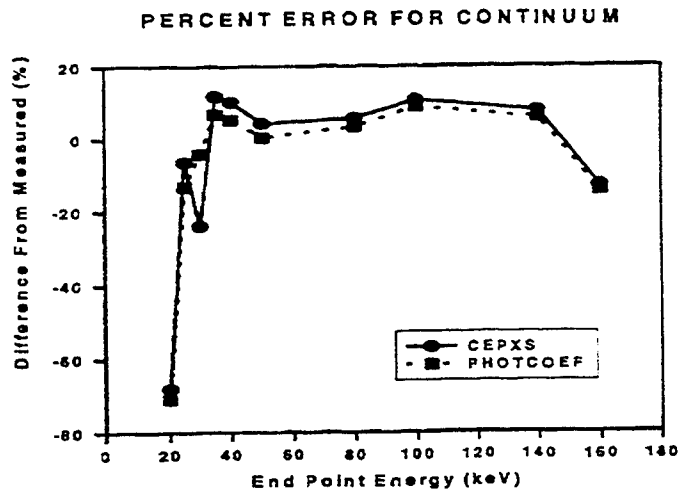


Figure 10. Error comparisons for calculations of continuum.

The L-lines contains the most ionizing energy for the entire spectrum when evaluating absorption in silicon. Figure 11 shows the percent error for the L-lines. The errors become very large as the anode voltage increases.

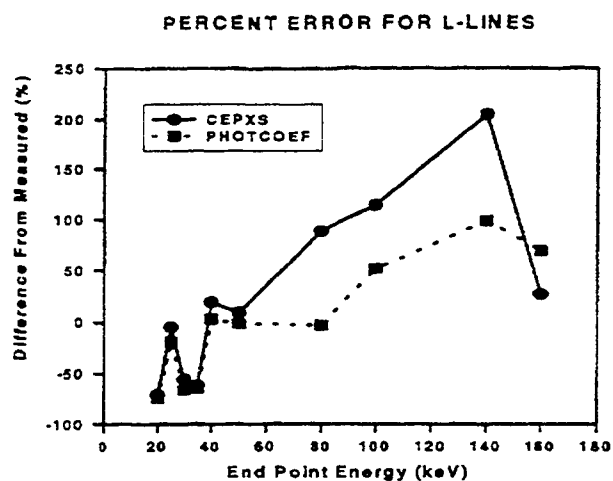


Figure 11. Error calculations for L-lines.

K-alpha and K-beta line data is shown in Fig. 12. The change in error from 140 to 160 kV is very unsettling and cannot be explained.

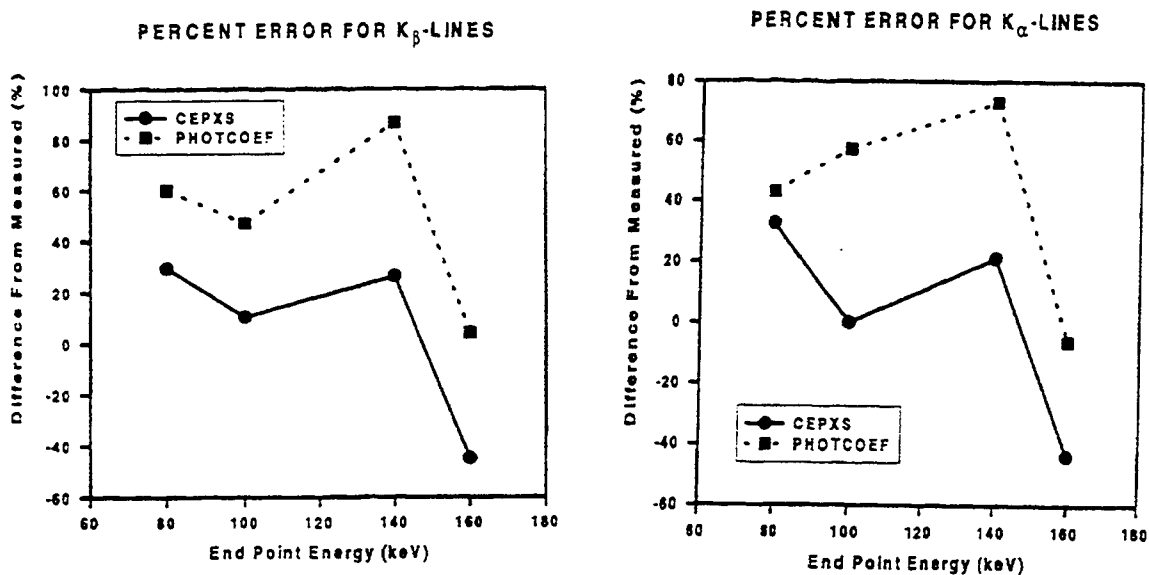


Figure 12. Error calculations for K-lines.

Part II

Characterization of CMOS Devices and Circuits

The trend in commercial MOS devices is toward smaller line geometries and a proportional scaling of most device dimensions and processes. As devices become smaller, the gate oxide thickness becomes very-thin ($<20\text{nm}$). As the oxide thickness decreases, the amount of interface-trap and oxide-trap charge decreases with slightly less than a $(t_{ox})^2$ relationship. Thus, the total-dose hardness of commercial gate oxides should improve as the gate oxide thickness is decreased.

Field oxides of advanced commercial technologies may however be very soft to ionizing radiation. Field oxides are much thicker than gate oxides, ranging from 200nm to 1000nm . Unlike gate oxides, field oxides are produced using a wide variety of deposition techniques such as CVD. Thus, the trapping properties of a field oxide may be poorly controlled and can be considerably different than for a thermally grown gate oxide. As radiation induced oxide charge builds up in a field oxide, it causes the threshold voltage of the field oxide to tend to go toward depletion mode for field oxides over a p substrate (equivalent to an n-channel field oxide transistor). If the buildup of charge is large enough, excessive leakage current can flow from the source to drain of the gate-oxide transistors and between transistors. The field-oxide leakage significantly adds to the drain-to-source current at zero gate voltage. Thus, the field-oxide leakage prevents the transistor from being completely turned off. This will greatly add to the static supply leakage current of an IC. Field oxide leakage current limits the radiation hardness of most commercial integrated circuits and it is a major problem for advanced hardened technologies, both at high and low dose rates.

This study evaluates the radiation tolerance of test chips fabricated using radiation hardening design rules. A commercial CMOS fabrication process is used with 1.2 micron line technology.

TEST SAMPLES AND EXPERIMENTAL DETAILS

Test Samples

Parts used in this work were designed and developed under a Phillips Laboratory contract with the Mission Research Corp. in Albuquerque. Radiation tolerant design techniques were incorporated where possible without affecting the IC processing steps. The layout rules are for the standard MOSIS 1.2 micron MOSIS process. The Hewlett Packard facility in Corvallis, Oregon was used for fabrication. The test chip incorporates the building block devices and circuits which are representative of those found in a high speed signal processor. The chip can be used to study a variety of radiation environments: total dose ionizing; single event effects (SEE); high dose rate upset and latch up. This study will only look at total dose ionizing. The 1.25 micron test structure contained the following test structures: a conventional two edge N-channel FET, a reentrant (edgeless) N-channel FET, see Figure 13, a two edge P-channel FET, N-channel with field oxide used as gate oxide, a single delay chain which is constructed of 19 NOR gates, and a shift register. The 1.25 test chip provided an opportunity to model the performance of the basic transistors which aided in the design of the 0.8 micron test chip.

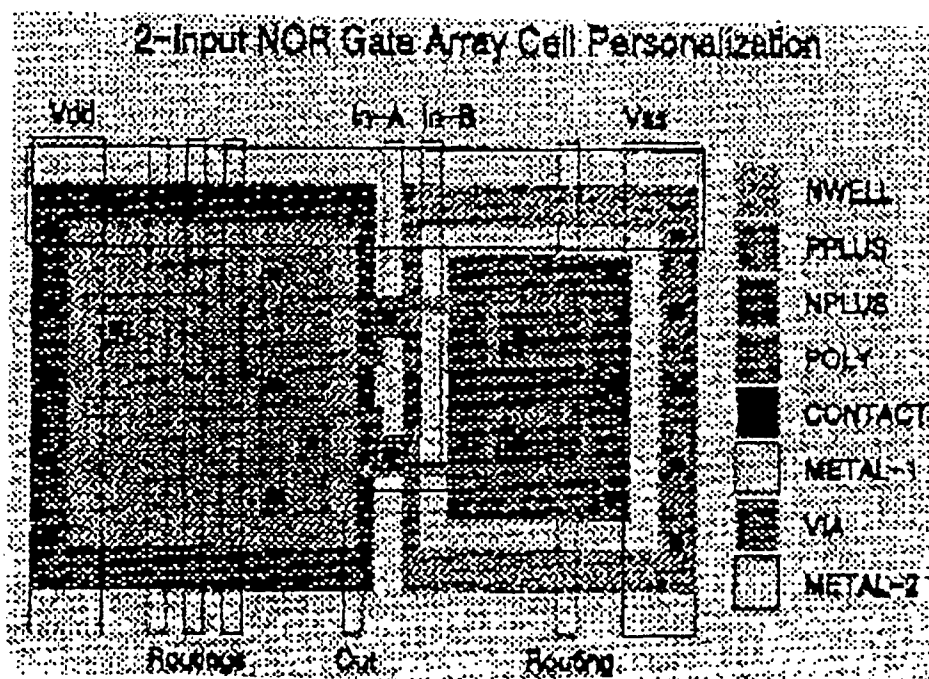


Figure 13. Two Input NOR Gate with Reentrant N-Channel FETs (Right Side).

The design goals for the 0.8 micron test chip were as follows: total dose to exceed 100 krad(Si), high single event upset immunity, and high dose rate hardness. The test chip was designed to measure and model the total dose performance of the process. The test structures chosen for the evaluation of this process were: a reentrant N-channel FET, a conventional two edge N-channel FET, a four edged P-channel FET sized to match the drive of the N-channel devices, two delay chains, one constructed of 101, three input NAND gates while the other made was 101, three input NOR gates, and N-channel FETs with field-oxides used as gate oxides. These circuits were designed to operate in a post radiation environment, independent of foundry.

MOSIS Processing

The 1.25 micron test chip was laid out with MAGIC version 6.0. The design rule checking software DRACULA was also used in the layout of the 1.25 micron test chip. The 0.8 micron test chip was laid out with the MAGIC design tool version 6.4.4. Other software tools used were buss routing BUSNET and Pre-Route Placement (PNC-PDSI). The HP MOSIS process uses a standard CMOS process with a thermal gate process and a LOCOS field oxidation process. The structure is an n-well process with a p-p+ epi starting material. The gate oxide (GOX) has a nominal thickness of 182 angstroms. The field oxide thickness is estimated to be 5-8000 angstroms. Both poly and Al gate processes were used. The metal interconnect pattern is three-level. Devices were bonded out in a 40 pin DIP. The lids were not sealed in order to allow radiation exposure and failure analysis. Anti-static handling techniques were strictly adhered to in order to prevent device damage due to static electricity.

Experimental Set-up

This experiment addresses the effects of total ionizing dose on the performance of test structures for a 1.25 micron and 0.8 micron process. Test data will not be shown for the 0.8 micron devices. All the test structures of interest on each chip were characterized before the devices were exposed to ionizing radiation. The characterization of all the FETs were taken with a HP-4145 semiconductor parameter analyzer.

The devices were irradiated in the Phillips Laboratory, Low Energy X-ray (LEXR) facility described in [1]. Irradiations were performed with the lids off the devices to minimize any effects of radiation attenuation or dose enhancement. The devices were irradiated at points within the facility that had been well calibrated and correlated to

Co exposure. A dose rate of 6.3 krad(Si)/min was used throughout all the irradiation's. Table 1 shows the total dose profile each device received between each characterization.

The devices reported were biased in their worse case condition during irradiation and anneal. After each exposure, the circuits were once again characterized. This process was repeated for each subsequent device. After the final exposure and characterization, the parts were annealed at 100 C for 168 hours in accordance with ASTM Test Method 1019.4. After completing the anneal, the parts were characterized one final time.

Total Dose Required	Exposure Time	Total DOSE/ Exposure
Pre rad	0.0 seconds	0.0 rad
10 krad(Si)	1 min, 32 sec	10 krad
20 krad(Si)	1 min, 32 sec	10 krad
40 krad(Si)	3 min, 5 sec	20 krad
80 krad(Si)	6 min, 10 sec	40 krad
160 krad(Si)	12 min, 20 sec	80 krad
320 krad(Si)	12 min, 20 sec	160 krad
480 krad(Si)	12 min, 20 sec	160 krad
640 krad(Si)	12 min, 20 sec	160 krad

Table 1. Radiation Exposures.

Results (1.25 Micron Devices)

While characterizations of the 1.25 micron devices/ circuit were carried out at total dose values listed in Table I, significant changes in performance were not detected for radiation exposures below 150 krad(Si).

By 150 krad(Si) measurable differences were observed, with the greatest change occurring in the conventional N-channel FET. Figure 14 shows the effects of 150 krad(Si) ionizing radiation on the reentrant N-channel FET, while Figure 15 shows the effects of 150 krad(Si) ionizing radiation on the conventional N-channel device. After irradiation the conventional N-channel FET exhibits substantial edge leakage, which is almost eliminated by the anneal.

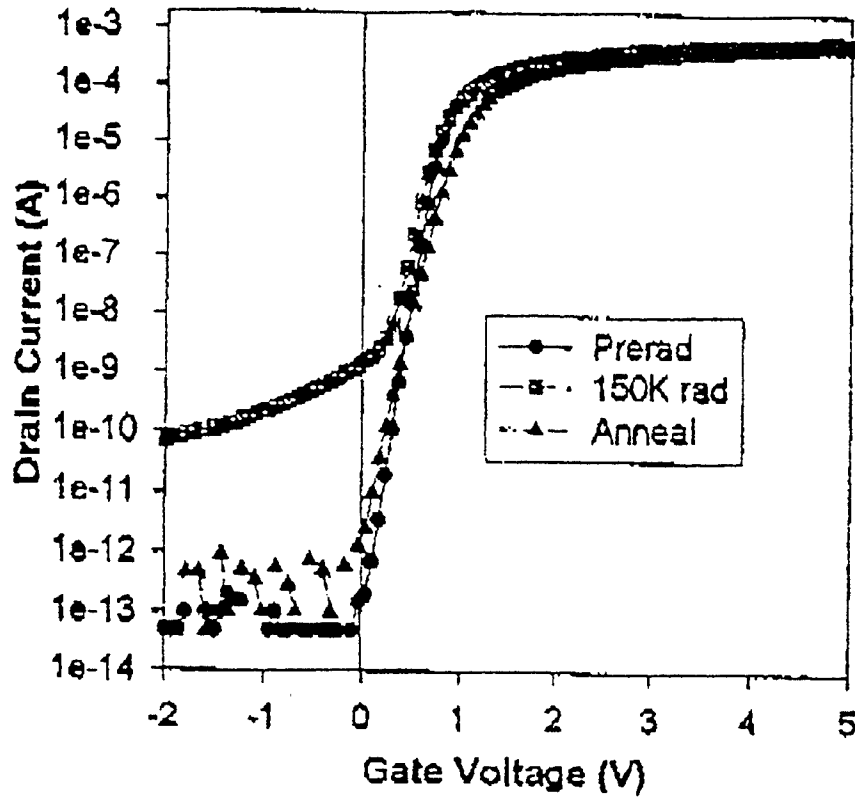


Figure 14. Subthreshold Characteristic of 1.25 Micron Conventional N-Channel FET after 150 krad(Si).

The leakage increased above the pre-irradiation leakage by approximately two orders of magnitude. After the anneal the edge leakage is reduced to near the pre-irradiation values. The threshold voltage shift is minimal both before and after the anneal has been performed.

The reentrant N-channel FET was then measured. Figure 15 shows the lack of edge leakage in the reentrant N-channel transistor and only a minor increase in the post irradiation total dose current leakage. In both transistors only very minor changes in threshold voltage are observed.

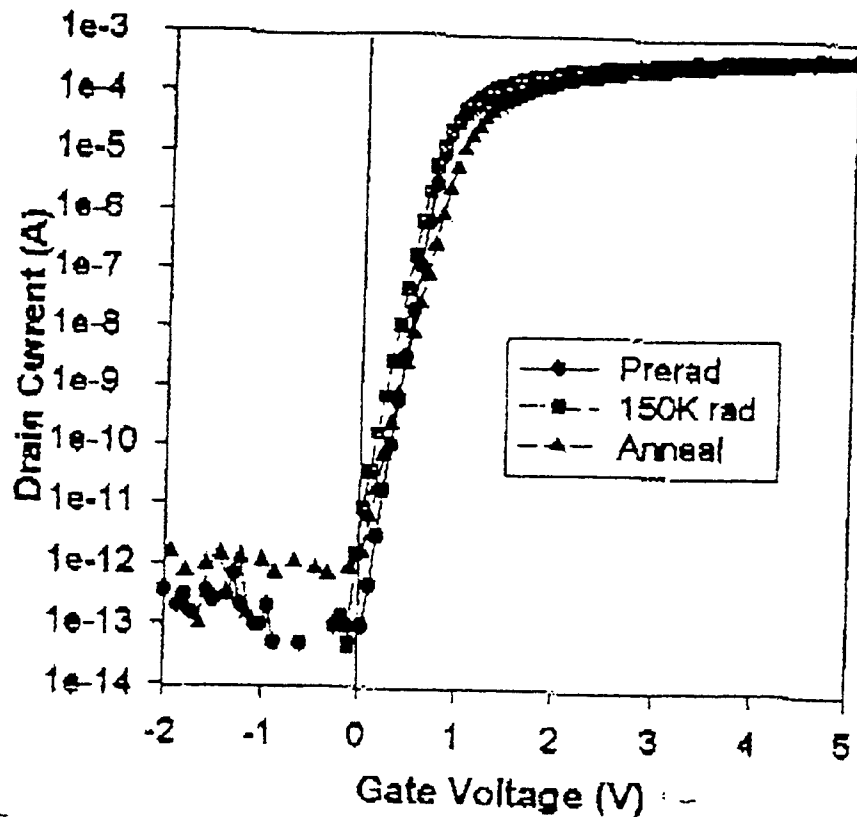


Figure 15. Subthreshold Characteristics of the 1.25 Micron Reentrant N-Channel FET after 150 krad(Si).

When the circuits are exposed to ionizing total dose radiation of 640 krad(Si) the edge leakage in the conventional N-channel transistors increases substantially as shown in Figure 17. After anneal the edge leakage decreased by two orders of magnitude but remains substantially above the pre-irradiation value.

The threshold voltage shifts substantially after anneal as the holes are annealed out of the gate oxide leaving the interface states which cause the threshold voltage to shift in the positive direction as shown in Figure 15, 16. In the reentrant N-channel FET the post-irradiation leakage increased only slightly, see Figure 16. The threshold voltage shift becomes a major portion of the change.

Immediately after the irradiation the threshold voltage shifted in the negative direction from the holes and the subthreshold slope decreased from the interface states, thus the FETs are not completely turned off with a gate voltage of zero volts, indicating that increased circuit leakage will be observed in actual circuits. After annealing the holes out of the oxide, the threshold voltage shifts back positive, but the subthreshold slope remains degraded. The reentrant N-channel FET degradation is much smaller than the degradation observed in the conventional FET design.

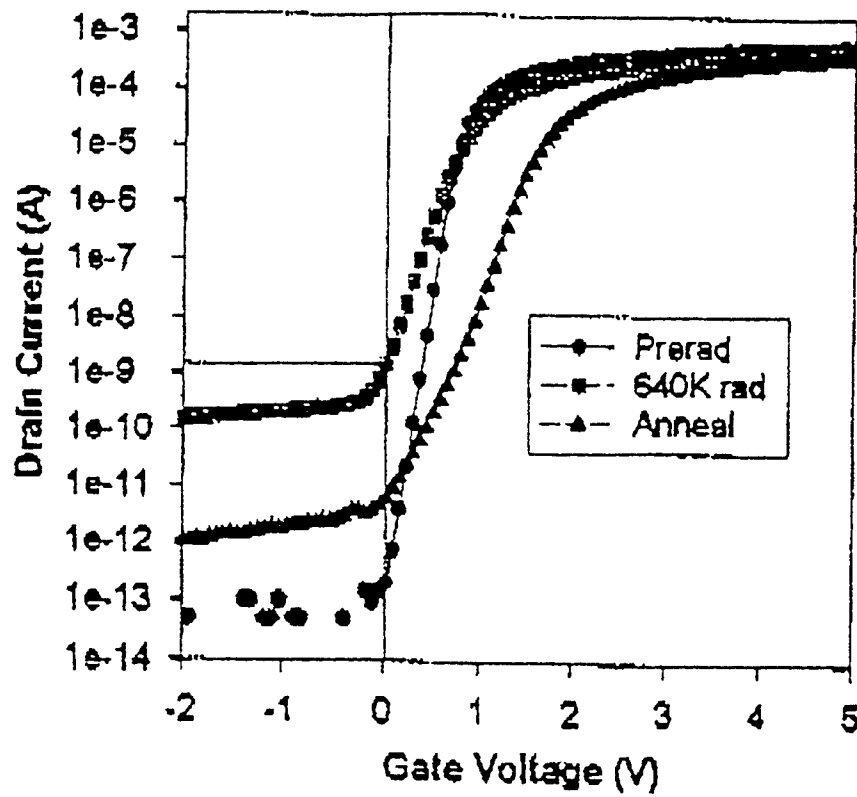


Figure 16. Subthreshold Characteristic 1.25 Micron Conventional N-Channel FET after 640 krad(Si).

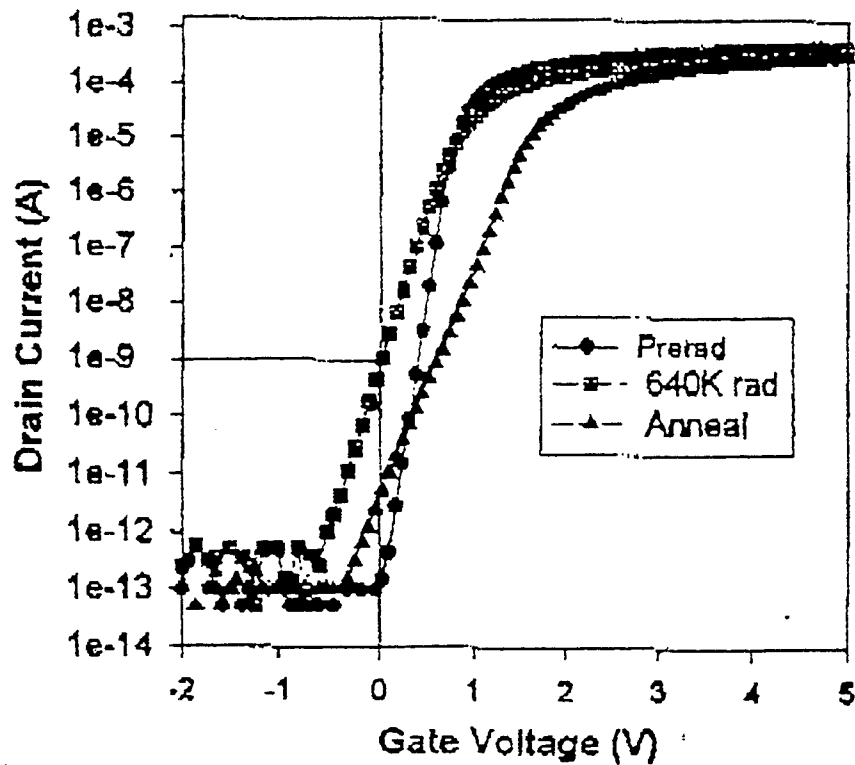


Figure 17. Subthreshold Characteristic 1.25 Micron Reentrant N-Channel FET after 640 krad(Si).

The P-channel transistors used in the 1.25 micron test chip are of a conventional design, see Figure 11. The subthreshold current plot for a P-channel FET exposed at 150 krad(Si) is shown in Figure 18. The increase in the off-state leakage current is approximately an order of magnitude larger than the pre-irradiation current, Figure 18. The leakage current values return to near pre-irradiation values after the 100 C, 168 hour anneal. The threshold voltage decreased by about 0.2 volts after exposure to the 150 krad(Si) radiation. It then recovered by 50% to the pre-irradiation value after the anneal.

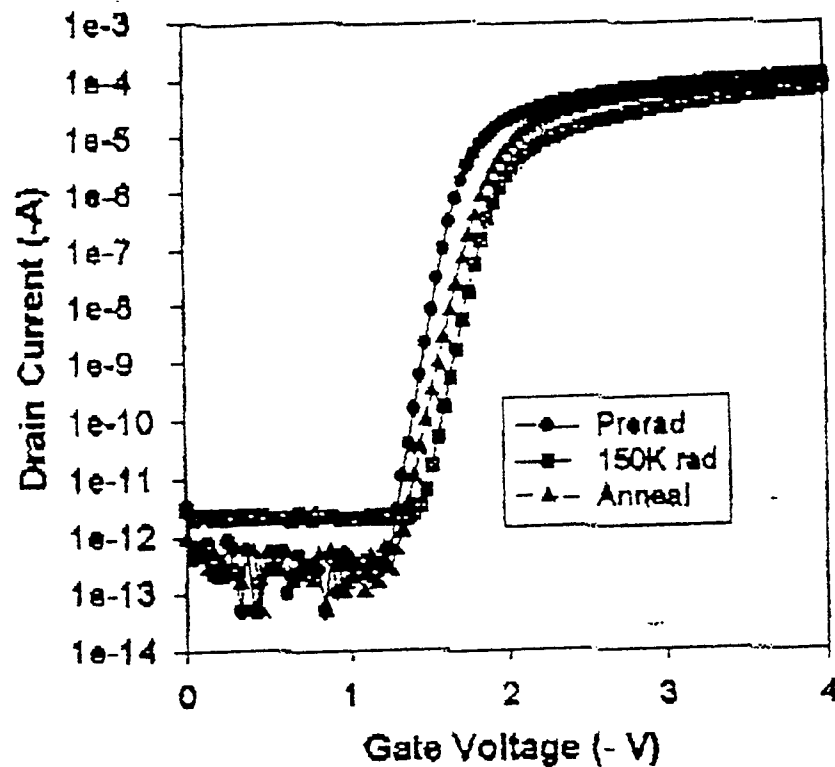


Figure 18. Subthreshold Characteristic 1.25 Micron Conventional P-Channel FET after 150 krad(Si).

A separate device was then irradiated to 640 krad(Si) and evaluated. The results of this evaluation are shown in Figure 19. After irradiation, the P-channel transistor performed very poorly, with both "off state" leakage increased and "on state" drive decreased. After annealing the current drive and leakage returns to near pre-irradiation values.

The P-channel threshold voltage is reduced by about 0.6 volts after exposure to 640 krad(Si), but returns to within about 0.2 volts of the pre-irradiation value after the anneal.

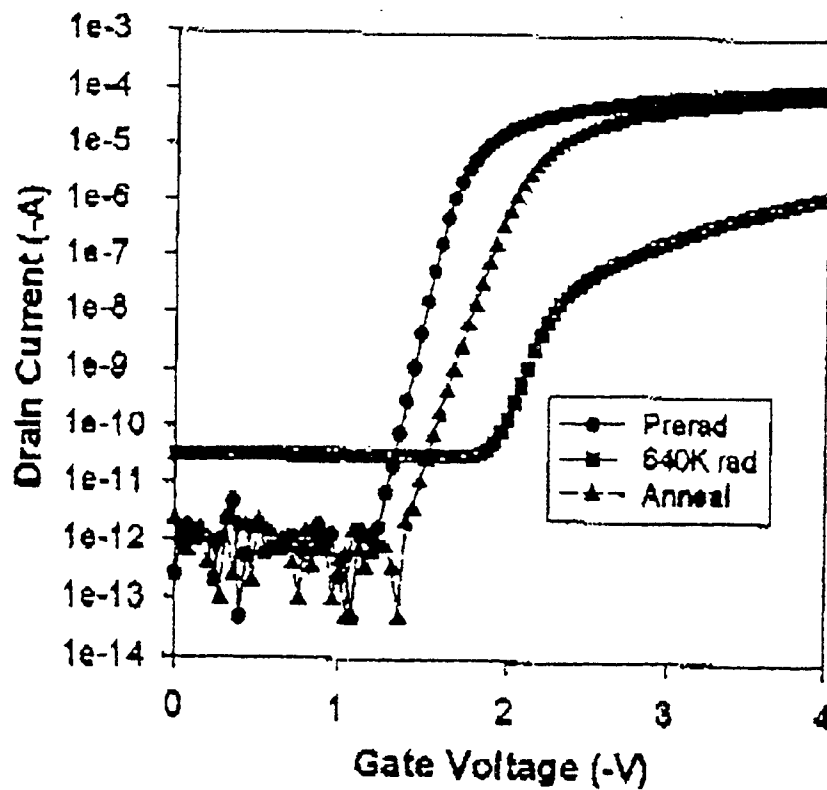


Figure 19. Subthreshold Characteristic 1.25 Micron Conventional P-Channel FET after 640 krad(Si).

The 19 NOR gate delay chains were evaluated. Each of the NOR gate used are of the configuration shown in Figure 11. In this evaluation the time delay is plotted with respect to the total ionizing dose exposure. Figure 20.

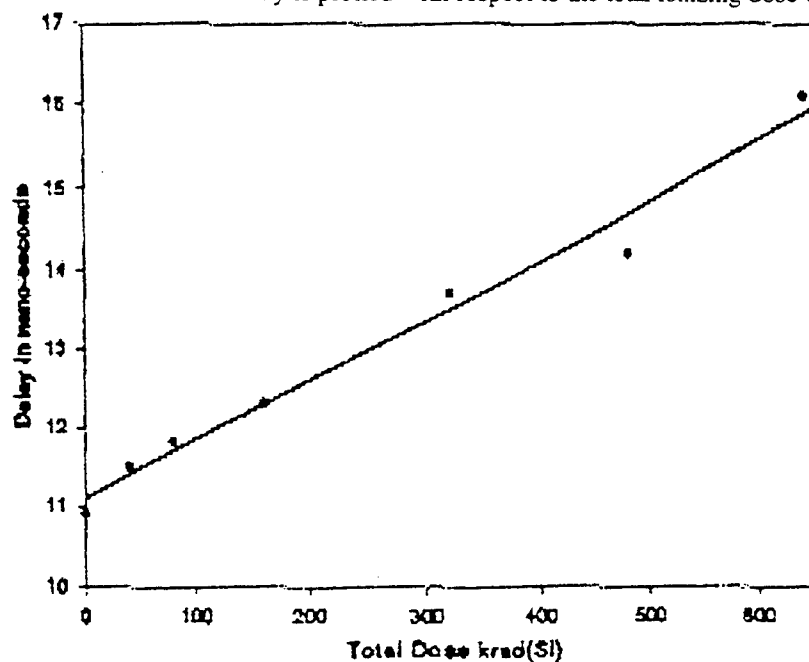


Figure 20. Delay Line Response Dose Vs Delay.

As plotted the delay time versus radiation exposure appears to increase in a linear fashion. This cannot be extrapolated to higher total dose levels due to the rapid degradation of the P-channel FET observed at 640 krad(Si). After the 640 krad(Si) exposure, the circuit was annealed and the results are shown Figure 21.

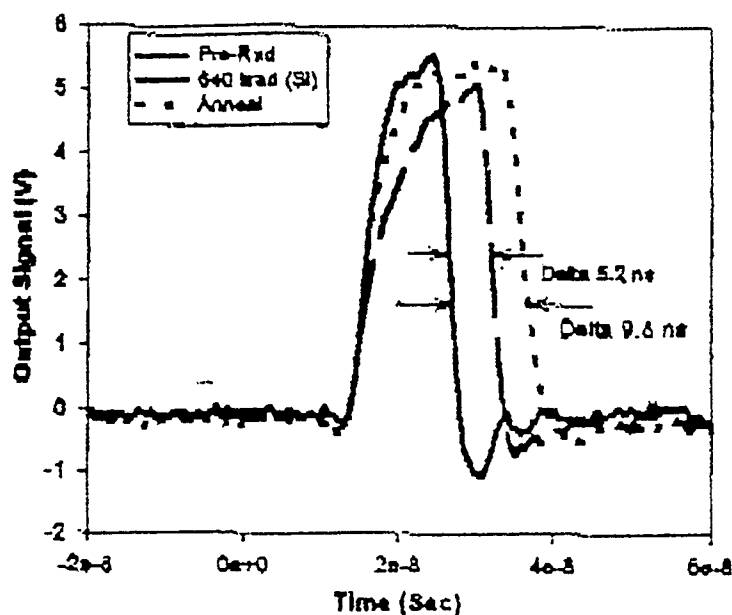


Figure 21. Delay Line Response Before and After Exposure to 640 krad(Si) and Anneal.

The increase in delay time observed after the anneal is due to the positive threshold voltage shift of the N-channel transistors in the gates, slowing down the turn on time of these transistors. It should be noted that while performance decreased after anneal, the static current of the transistors also decreases.

Conclusion

Part I

This effort has been on-going now for three years. While we have had a significant number of good publications over the radiation characterization of the LEXR facility, there are several tasks to complete in order to wrap-up this part the project. Additional silicon transmission data should be taken to more accurately define the correlation between the measured and calculated transmission data. Tube current monitoring is critical in this measurement and has been a problem in the past. Furthermore, the HPGe detector should be modeled for both detector response and a 3D model to include the detector and enclosure to reveal possible anomalies in the very low energy region ($>5\text{KeV}$). With this additional work, the author is of the opinion that a comprehensive paper could be written covering the entire block of work which would answer questions regarding discrepancies found by Beutler & Halbleib [5] in 1989.

Part II

The characterization of MOSIS parts began during the Summer of 1994. Data has been taken and analyzed on both 1.25 micron devices and on 0.8 micron devices. There are a number of projects which lend themselves to further research in this area. Perhaps the most intriguing is the simulation of the processing, device characterization and circuit response using the Silvaco Tool Set. Model enhancements can be developed which will enhance the tool set so that simulation of total dose and eventually dose rate effects can be accomplished in MOS devices. Transient and SEU studies can also be accomplished with devices and circuits fabricated using both geometries. Flash-xray testing can characterize both transient upset and latchup susceptibility.

References

- [1] X-ray Wafer Probe for Total Dose Testing, IEEE Trans.on Nuclear Sci., NS-29, Dec. 1982.

- [2] X-Ray Spectroscopy Measurements of a Low Energy X-Ray Facility,"; IEEE Transactions on Nuclear Science; Vol.41, No.6; pp.2112-2117, Dec. '94

- [3] Calibration of a Low Energy X-Ray Facility; "IEEE Transactions on Nuclear Science; Vol. 41, No. 6; December 1994

- [4] Application of Analytical and Experimental Dosimetry Techniques to a Low-Energy X-Ray Radiation Source;" Journal of Radiation Effects; Vol. 12, No. 3; pp. 28-35; Sept. '94

- [5] "Comparison of Experimental Pulse-Height Distributions in Ge Detectors With Integrated TIGER Series Code Predictions," IEEE, Trans. Nucl. Sci, Vol 36, #6, Dec 1989

Jan Brzosko report unavailable at time of publication.

PARALLEL COMPUTATION OF ZERNIKE ABERRATION COEFFICIENTS FOR
OPTICAL ABERRATION CORRECTION

Meledath Damodaran
Professor
Division of Arts & Sciences

University of Houston - Victoria
Victoria, TX 77901

Final Report for:
Summer Research Extension Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

University of Houston - Victoria

December 1995

PARALLEL COMPUTATION OF ZERNIKE ABERRATION COEFFICIENTS FOR OPTICAL ABERRATION CORRECTION

Meledath Damodaran
Professor

Abstract

Aberrations occur in images in every optical system. In order to correct for these optical aberrations, they must be detected and characterized. The aberrations are characterized by using orthogonal basis functions composed of discretized Zernike polynomials. The coefficients associated with each Zernike polynomial can be measured using a Phase Diversity wave front sensing technique. Nonlinear optimization techniques are utilized to calculate the Zernike coefficients in a serial manner. We showed that each Zernike coefficient can be calculated independently in a parallel fashion from each other and the same results are realized. We describe here a parallel implementation of our algorithm using the IBM SP2 parallel computer. We used the PVM software for parallelizing the computational tasks across the processors in a "master/slave" fashion. The computation can be performed in an efficient manner using this strategy. Concurrent evaluation of Zernike coefficients also has application in their evaluation using neural networks. Here we describe a strategy to recursively compute the Zernike coefficients using a method similar to cascade correlation.

PARALLEL COMPUTATION OF ZERNIKE ABERRATION COEFFICIENTS FOR OPTICAL ABERRATION CORRECTION

Meledath Damodaran

Introduction

Computer scientists and engineers have been interested in understanding and then improving the process of image formation. Obtaining any improvement is especially challenging when something in the image formation process corrupts the image. This requires an understanding of the image formation process, including which processes degrade the image. Once an accurate understanding of the distorting processes is acquired, the researcher has an opportunity to use image reconstruction and image deconvolution techniques to enhance the images. Therefore, it is obvious that the better the degrading process is characterized, the more success one will have in undoing the process to recover a more accurate image.

Image degradation occurs in a variety of imaging systems including photography, industrial radiography, microscopy, medical imaging, remote sensing, aerial reconnaissance, ground-based imaging and space-based imaging [6]. Fortunately, a few types of degradation phenomena appear repeatedly, therefore, good degradation models and systems have been developed. Descriptions of the models are needed to mathematically characterize the image formation process.

The presence of aberration in an optical system can be represented by Zernike polynomials [1]. The Zernike polynomials are usually described in terms of the radial variable and the angular variable in the pupil. For the purpose of correcting atmospheric aberrations it may be necessary to estimate several coefficients in the Zernike decomposition of the wave front function. In [2] a method using Phase Diversity to extract higher order Zernike coefficients was introduced. The method is based on non-linear optimization involving several variables.

The Phase Diversity method may be summarized as follows. Two images are acquired: one at the focal plane and the other with some known aberration (typically focus aberration). The first image is corrupted by the unknown aberration, whereas the second image (known as the diversity image) contains not only the unknown phase errors of the

system but also the additional diversity phase error. Using the optical transfer functions of the system with and without the added defocus, and the Fourier transform of an estimate of the object, an error metric is set up between the measured and diverse images. This error metric, originally due to Gonsalves [3], is independent of the original object estimate. An excellent source of derivation of the Gonsalves metric can be found in [6]. A nonlinear optimization method such as the gradient descent method is then used to estimate the Zernike coefficients that would minimize the above error metric. Carreras et al [2] used the conjugate gradient method for nonlinear optimization, using finite difference methods to calculate gradients.

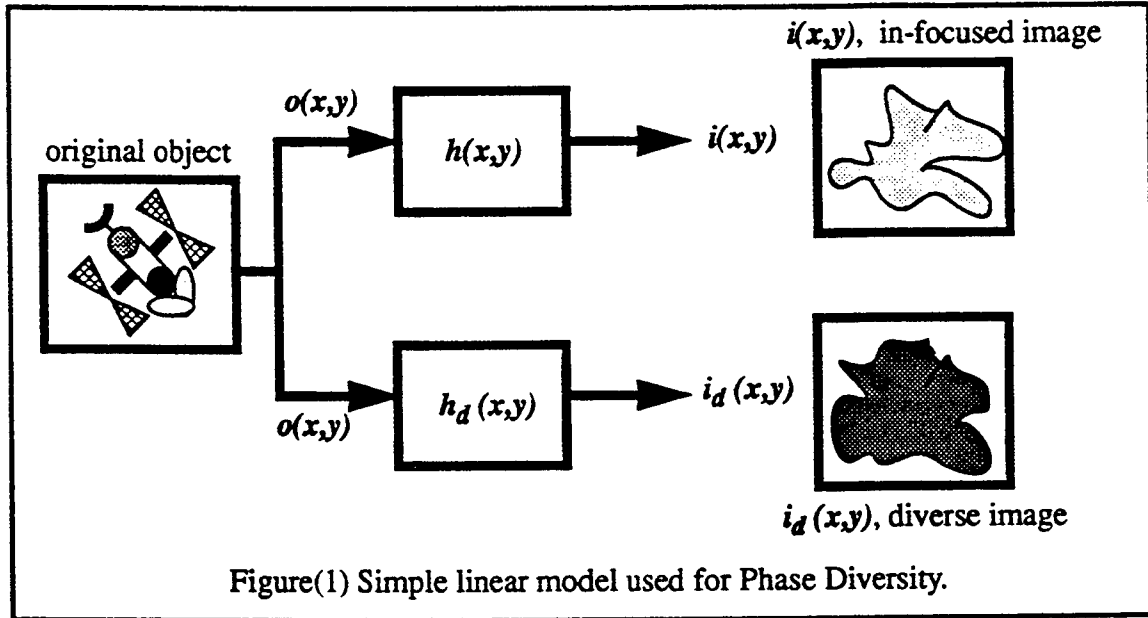
Even though the above method is attractive, it is computationally a very formidable task to calculate several Zernike coefficients for a given system. In order to get a reasonably accurate approximation of the coefficients, a large number of iterations had to be carried out. Each iteration involves time-consuming calculations of the error function and the gradients. These could not be done in real time, to correct images on the fly.

In an earlier paper we showed that all the Zernike coefficients may be computed independent of each other [9]. In this paper we introduced a method using non-linear optimization involving only a single variable to independently evaluate the different Zernike coefficients. Independent evaluation of the Zernike aberration coefficients implies the possibility of using a parallel processor or a network of processors to speed up the computation. In this paper we outline our approach to parallel evaluation of Zernike coefficients using the IBM SP2 Parallel Computer at the Maui High Performance Computing Facility in Maui, Hawaii, a machine owned by Phillips Lab. We used the IBM SP2 parallel computer with the PVM parallelizing software for this implementation. In this program each processor independently evaluates a Zernike coefficient under the control of a "master" program.

Concurrent evaluation of Zernike coefficients also has application in their evaluation using neural networks. Here we describe a strategy we used to recursively compute the Zernike coefficients using a method similar to cascade correlation.

Methodology

We start with a pupil function P in the form of a 64 by 64 matrix representing a circle. The circle itself is contained in a 32 by 32 submatrix. The imaging system used is as shown in Figure 1.



All variables shown in the figure are in spatial domain. The original image can be seen to go in two directions, to the in-focus optical system and to the out-of-focus or diverse optical system. The functions $h(x,y)$ and $h_d(x,y)$ are the point spread functions (PSF) for the two systems, and $i(x,y)$ and $i_d(x,y)$ are the two received images in the two systems. We then compute the generalized pupil function P

$$P = A(u,v)e^{iF(u,v)},$$

where $A(u,v)$ is the pupil function of the optical system and $F(u,v)$ is the basis function used in optics to describe the various aberrations and is composed of discretized Zernike polynomials:

$$F(u,v) = a_1 Z_1 + a_2 Z_2 + \dots$$

Here Z_1, Z_2 etc. are the Zernike polynomials and the coefficients a_1, a_2 etc are the Zernike aberration coefficients, and

represent piston, tilt, focus etc. In our simulation, the first three Zernike aberrations were not used, as they are not pertinent to our application. However, our method may be applied with any number of aberrations. The terms a_4, \dots, a_8 were given values representing simulated aberrations, and represent focus error, astigmatism in the x-direction, astigmatism in the y-direction, coma in the x-direction, and coma in the y-direction respectively. We compute the optical transfer function H by convolving P with itself. Similarly we compute the generalized pupil function for the diverse image as follows:

$$P_d = P \exp(ikZ_4)$$

and the corresponding transfer function H_d by convolution. Here, k is the coefficient of defocus in the diversity term.

Next, a computer-generated test object was chosen. This pristine object is also of size 64 by 64 pixel elements. The object is next blurred using the transfer function $H(u,v)$ to generate the in-focus measured image. In addition, an out-of-focus image is generated by using the transfer function $H_d(u,v)$. The input-output equations in the Fourier domains will be:

$$I(u,v) = O(u,v)H(u,v)$$

$$I_d(u,v) = O(u,v)H_d(u,v)$$

The non-linear optimization can start now. For each separate aberration we compute a P' and a P_d' as follows:

$$P' = P \exp(i a_j' Z_j)$$

$$P_d' = P' \exp(i k Z_4)$$

where $j = 4, \dots$, and $'$ denotes estimates. We then take the 2-D convolutions of these to get the estimated transfer functions H' and H_d' :

$$H' = P' * P'$$

$$H_d' = P_d' * P_d'$$

The Gonsalves error metric can now be calculated as:

$$E = \sum [|H_d' - I_d H'|^2] / [|H'|^2 + |H_d'|^2],$$

where the summation is over all the matrix elements.

The function E can be seen as a function of a_j' alone, and is independent of the object. It is, furthermore, a smooth

function. This is a generalization of the error metric first derived by Gonsalves [7,8].

In order to determine the value of a_j' for which E is minimum, we used a quadratic interpolation method due to M. J. D. Powell, from the IMSL library of mathematical subroutines [5]. We chose this routine because it uses only function evaluations and because it gave satisfactory results when we used double precision arithmetic. With single precision arithmetic, rounding errors in the calculations caused the existence of, and convergence to, other local minima nearby the sought-out minimum point.

Parallel Implementation of the Concurrent Phase Diversity Algorithm

The IBM SP2 machine at Maui High Performance Computing Center is a state-of-the-art MPP (Massively Parallel Processor) computing system owned by Phillips lab and managed under a cooperative arrangement by the Albuquerque Research Center of the University of New Mexico. It consists of a large collection (400 as of now) of IBM RS 6000 high performance chips connected by a high speed switch. Any processor in the system can communicate with any other processor via message passing. There is no shared memory and the system is equipped with an NFS mounted Common File System. The parallel system can be configured under user programming control into any shape or form. The machine can achieve Grand Challenge performance. We used the PVM software [10,11] to parallelize the code that we had earlier developed for the IBM RS 6000, so as to run on the SP2 machine. PVM is a popular software system allowing the development of large distributed parallel applications, supporting the message passing mode of parallel computation. The PVM application was created using the system calls of the PVM Library.

The goal of this part of the project was to convert the phase diversity program from single processing to parallel processing program using PVM to reduce the calculation time by increasing the number of parallelly executable tasks while at the same time keeping the computation-to-communication ratio relatively high. The first step was to understand how PVM is implemented in programs. The second step was to select the proper implementation of PVM that best suited the program. The last step was to test the program for speed increase.

PVM environment is quite simply a network of computers, homogeneous or not, running the PVM daemon

process. The daemon process is interfaced using a PVM library for use with C or Fortran which would allow PVM programs to send messages between themselves on the network. Because the original program was implemented in FORTRAN, and would require much time to convert to a C counterpart, we used the FORTRAN APIs. The implementation we chose was to separate the five large contributors of processing time in the program and distribute the load across the processors. The format we chose was the master/slave model versus the single program multiple data (SPMD) model because of the nature of the program we were converting. Because of library calls made in the single processor version, the parallel version would require that the master and each slave would need to have access to the libraries that were used in the original version.

After reviewing the initial attempts of the parallel code extensively, we found that the pieces of code we had distributed were not the largest contributors in time but very small contributors in time. With his help we re-selected the functions that were contributing the most time and rewrote the program to properly execute the distributed processes. To do this we moved several functions to the slave processes.

The main algorithm task that was distributed across different physical processors was the concurrent nonlinear optimization step resulting in the computation of each Zernike coefficient. We parallelized across the number of optimizations performed. The master program spawns as many child processes as the number of Zernike coefficients to be calculated, and waits for each child to complete its task and send the result back to it. Since the granularity of the computation is at a high level, this strategy is very attractive for raw speedup on a network computer such as the SP2.

The code was compiled and tested thoroughly with the IMSL library calls commented out because we could not find a site that had both the PVM and the IMSL libraries necessary to run the program. In studying the code we found that limitations in the mathematical libraries caused the program to be slightly less robust. Because the library subroutine DUVMIIF required a single parametered function as input we had to actually have five sets of code for slave processes. Had the subroutine allowed more than one parameter we could have had only one set of code for the slave processes and used a massively-parallel-processor (MPP)- like approach. Because the number of processes were maximized at five we anticipate a somewhat linear speed increase over the single processor implementation. If the optimization and the FFT

processes were not in a library the granularity of the parallelization could be greater, but we expect that communication overhead (in interprocess communication) would then start having a significant impact on the overall execution. We feel that the implementation that we have created now is the right balance to achieve the optimum speedup. We feel confident that the timing runs (that we expect to be able to conduct in January after IMSL has been loaded at the MHPCC facility, would prove this claim of ours.

Parallel Computation of Neural Networks using Independent Back Propagation Neural Networks

In order to compute several aberration coefficients, thanks to the independence of these coefficients, we may use as many number of simpler back propagation neural networks rather than one complicated network that simultaneously computes all the coefficients. This was our general strategy as we tried to apply neural network technology to solve the problem of high speed concurrent computation of Zernike coefficients. The attractive feature of neural networks is that even though the training time is relatively large, once the networks are fully trained, it is a very fast computational engine.

We tried increasingly more sophisticated strategies to solve this problem. But in all cases we used point source objects. The optical transfer functions H and HD were computed using given values of the five aberration coefficients. We then computed a metric

$$M = (H - HD) / (H + HD) \quad (1)$$

The advantage of this metric is that it uses both the in focus and defocus optical transfer functions, and have been successfully used by other researchers.

We reduced the size of this 64 X 64 matrix M to 16 real numbers by performing repeated local averaging. These 16 values and the corresponding aberration coefficients were used as the input-output pair for training and testing of the back propagation neural networks that formed the cascade network. We repeated this for each set of 3000 previously created simulated aberration coefficients, thus constructing a total training and test data set of size 3000. Each neural network had one hidden layer of 10 neurons and one output neuron. The input layer consisted of 16 neurons (not

including one for the bias) for the first network, 17 neurons for the second, etc. The last network used 21 input layer neurons. The 16 real numbers were input into the first network, which computed the first aberration coefficient; the same 16 inputs along with the first computed aberration coefficient were input to the second network, which computed the first aberration coefficient; these, along with the second computed aberration coefficient were input to the third network; etc. in a cascaded fashion. We used zero momentum and a relatively small learning rate of 0.02.

Results.

Sequential Implementation of the Concurrent Phase Diversity Algorithm

An average of 12 iterations was all that was required for convergence in the five instances of single variable optimizations of our experiment. The average percentage error of the computed values a_j' from the given values a_j was no more than 15 percent. This error may be attributed to the effect of round-off errors, and is unavoidable in minimization routines that only use function values. The results are summarized for a typical computer run with an extended object in Table 1. Table 1 shows the number of the Zernike coefficient in the first column. the second column shows the exact value of the coefficient which is inserted as the simulated aberration. The third column shows the value which the algorithm computationally estimated. This technique should be evaluated by comparing the estimated value in column 4 and the actual value in column 3. Finally the values of the Gonsalves error metric to which the algorithm converged, are shown in the last column. The difference between actual values and estimated values are no more than similar runs in the multivariate case [2] and are well within expected norms. In fact, the image reconstruction with this level of accuracy is quite satisfactory, as illustrated in [2]. Explicit derivative evaluation is not practical, as the dependence of E on a_j' is via a convolution; hence, optimization routines that explicitly require the derivative to be made available to the routine, cannot be employed.

However, when we ran the simulation using a point spread image instead of an actual image, the estimated Zernike coefficients were near perfect; the average error in this case was just 1.7 percent.

TABLE 1 Simulation Results

<u>Zernike Number</u>	<u>Actual Coeff. Value</u>	<u>Estimated Coeff. Value</u>	<u>Error Metric</u>
Z ₄	-1.25	-1.23	0.1e-4
Z ₅	0.66	0.78	0.7e-4
Z ₆	1.5	1.74	0.66e-4
Z ₇	0.44	0.43	0.77e-4
Z ₈	-0.33	-0.21	0.77e-4

Average percentage error : 14.7

Parallel Implementation of the Concurrent Phase Diversity Algorithm

The parallel code was compiled and tested thoroughly with the IMSL library calls commented out because we could not find a site that had both the PVM and the IMSL libraries necessary to run the program. Maui High Performance Computer Center now has a copy of IMSL libraries and they have promised to have this installed soon. So, we only can report on the theoretical speedups we expect to get from our implementation at this time. Because the number of processes were maximized at five we anticipate a somewhat linear speed increase over the single processor implementation. If the optimization and the FFT processes were not in a library the granularity of the parallelization could be greater, but we expect that communication overhead (in interprocess communication) would then start having a significant impact on the overall execution. We feel that the implementation that we have created now is the right balance to achieve the optimum speedup. We feel confident that the timing runs (that we expect to be able to conduct in January after IMSL has been loaded at the MHPCC facility, as promised by them) would prove this correct.

Parallel Computation of Zernike Coefficients Using Independent Back Propagation Neural Networks

The cascade neural network was set up using five back propagation neural networks. The first network (ZER_A1) used as its inputs the sixteen data fields generated from the metric M of equation (1) after applying local averaging repeatedly. The other networks were built by cascading them to the previous network's output. Thus, each other network used the same original sixteen data fields and the output(s) of the previous networks as its inputs. Each of the five networks was built having those inputs in the input layer, ten processing elements in one hidden layer, and one output in the output layer.

The learning rate used in each network was set at 0.02. The momentum was initialized at 0.00 for each network. The learning rate and the momentum did not change during any part of the training of the networks. They applied to each layer of the network and used these values throughout the complete training. In the output layer and the hidden layer, the sigmoid function was assigned as the transfer function and the delta rule was designated as the learning rule. The parameters were chosen for the networks so as to provide the best results during both training and testing.

Conclusion

We were able to show that the estimation of the aberration coefficients may be done independent of each other using the same Gonsalves metric as in the normal case. That the accuracy of these estimates was comparable to, if not slightly better than, when they were calculated by the straightforward method of non-linear optimization involving several variables, demonstrates that the method of independent evaluation using a non-linear optimization of the Gonsalves metric with a single variable, is experimentally valid. It would be good to prove this theoretically as well. It is remarkable and somewhat surprising that when we consider only one aberration at a time in the calculation of the Gonsalves metric, the terms $I(x,y)$ and $I_d(x,y)$ were derived from contributions from all the aberrations, whereas the estimates $H(u,v)$ and $H_d(u,v)$ only have that one aberration under consideration, and still it does not seem to affect the minimum point.

One of the advantages of independent evaluation of the Zernike coefficients is that the total amount of time required to calculate all the coefficients is dramatically shorter than in the multivariate case, as (i) the total number of iterations needed is less, and, more importantly, (ii) each iteration in our algorithm takes only a fraction of the time for

the multivariate case, it being the case that it is easier for the function to move towards a minimum in the case of a function of a single variable. The fact that the calculation of the coefficients may be done independently, also means that we could partition the task into several nodes in an interconnected network of processors such as the SP2. Our successful parallel program shows that for computing a large number of aberration coefficients, of the order of 20 or 30, in a real time image reconstruction scheme, a scalable high speed network computer such as the IBM SP2 will be a suitable choice.

Another case in point is the application of neural networks. In order to compute n aberration coefficients, we may use n simpler neural networks rather than one complicated network that computes all the coefficients at the same time. In this case, the neural network implementation would also be a parallel one. It would then be possible to apply this method in a real-time image reconstruction scheme.

The first two networks, (ZER_A1 and ZER_A2), have been trained and tested, providing positive results in their outputs. During the training of the first network (ZER_A1), the root mean square (RMS) error showed convergence and approached zero. The classification rate, the percentage of the actual outputs produced when compared with the desired outputs, is 94.41 percent.

The second network (ZER_A2) was also trained and tested. The performance of this network was also positive in achieving good results. The RMS error in this network showed a slightly better result during the training phase. The network converged toward zero and was actually closer to zero than the first network. This second network, however, did not produce as good results as far as the classification rate is concerned. The value of the classification rate was 92.29 percent. This value is slightly lower than the previous network.

The final three networks have not yet produced results matching the first two. It is well known that neural network parameters and architecture have to be individually calibrated to the problem at hand, and we are continuing this work.

Rapid calculation of Zernike aberration coefficients has many different applications. It can be used in post-

processing of images, wave front sensing, and possible closed loop adaptive optics correction. It will be used to correct for aberrations in images from the various Air Force telescope observational sites.

References.

- [1] C-J. Kim and R.R. Shannon, "Catalog of Zernike Polynomials", Ch. 4, Applied Optics and Optical Eng., vol. X, pp193-221, edited by R.R. Shannon and J.C. Wyant, Academic Press
- [2] R.A. Carreras, S. Restaino, D. Duneman, " A Laboratory Experiment using Phase Diversity to Extract Higher Order Zernike Coefficients", Proc. of the SPIE Int'l Symp. on Optics, Imaging and Instrumentation, July 1994
- [3] R.A. Gonsalves, "Phase Retrieval and Diversity in Adaptive Optics", Optical Engineering, vol. 21, no. 5, Sept./Oct. 1982
- [4] T.K. Barrett and D.G. Sandler, "Artificial Neural Network for the Determination of Hubble Space Telescope Aberration from Stellar Images", Applied Optics, vol. 32, no. 10, April 1993
- [5] Users Manual, FORTRAN Subroutines for Mathematical Applications, vol.3, IMSL, Inc., Houston, TX, 1989
- [6] R.G. Paxman and J.R. Fienup, "Optical Misalignment Sensing and Image Reconstruction Using Phase Diversity", J. Opt. Soc. Am. A, vol.5, no. 6, June 1988
- [7] R.A. Gonsalves, "Phase Retrieval From Modulus Data", J. Opt. Soc. Am., vol. 66, no. 9, Sept. 1976
- [8] R. A. Gonsalves and R. Chidlaw, "Wavefront Sensing by Phase Retrieval", SPIE, vol. 207, Applns. of Digital Image Processing III, 1979
- [9] Richard A. Carreras, Greg Tarr, Sergio Restaino, Gary Loos, and Meledath Damodaran, "Concurrent Computation of Zernike Coefficients used in a Phase Diversity Algorithm for Optical Aberration Correction", Proceedings of the European Symposium on Satellite Remote Sensing, Rome Italy, September 1994
- [10] Al Geist et al., PVM 3 User's Guide and Reference Manual, Oak Ridge National Laboratory, May 1994.
- [11] Robert Manchek, "An Introduction to PVM (Parallel Virtual Machine)"; Aug 5, 1993.

QUALITY FACTOR EVALUATION OF COMPLEX CAVITIES

Ronald R. DeLyser
Assistant Professor
Department of Engineering

and

Peyman Ensaf
Graduate Student

University of Denver
Denver, CO 80208

Final Report for:
Summer Research Extension Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

University of Denver

December 1995

QUALITY FACTOR EVALUATION OF COMPLEX CAVITIES

Ronald R. DeLyser
Assistant Professor of Electrical Engineering
Peyman Ensaf
Graduate Student
Department of Engineering
University of Denver

ABSTRACT

This paper investigates the electromagnetic field distribution within a Celestron-8 Cassegrainian telescope. This telescope, in general, represents a complex cavity which is a metal enclosure which has no known analytical solution for the electromagnetic fields. Such cavities quite often contain sensitive elements which may be susceptible to electromagnetic radiation. Therefore, it is important to be able to calculate or to measure the electromagnetic energy levels that these elements may experience and to determine potential performance degradation. This report presents some of the numerical techniques utilized for this analysis and compares these results with measurements. It also addresses some of the problems and intricacies encountered while conducting the analysis and details some of the steps taken to rectify these problems, offering guidelines for further analysis.

QUALITY FACTOR EVALUATION OF COMPLEX CAVITIES

Ronald R. DeLyser
Peyman Ensaf

Background

The objective of this project was to perform numerical analyses on a Celestron-8 telescope (Figure 1) to determine its susceptibility to high power microwaves. In order to analyze the telescope at low frequencies, our initial approach was to model the telescope's outer most shell as an open ended cylinder with an end cap at the top. Even though an aperture for the eyepiece resides at the center of the top of the telescope, we decided to model this end of the telescope as a closed surface. This was done because the aperture for the eyepiece is very small at the lower frequencies of interest. We modeled the resulting open ended cylinder using Mathematica [1], and analyzed that model using CARLOS-3D [2] (Code for Analysis of Radiators on Lossy Surfaces) over a frequency range of 0.7 GHz - 3.0. We then proceeded to create a model which included the interior geometry of the telescope which consisted of two interior cylinders and the aperture for the eyepiece. The interior cylinders (Figure 1) are at the center of the large aperture, where a convex reflecting mirror resides (the short plastic cylinder) and at the eyepiece (the long metal cylinder). The two reflecting mirrors are not shown in Figure 1. The CARLOS-3D model (Figure 2) was analyzed over a frequency range of 0.5 GHz to 3.0 GHz. Each of the models, the open cylinder and that shown in Figure 2, are discretized with triangular cells, each having sides no longer than $1/10$ of a wavelength (λ) in length corresponding to a frequency of 1.0 GHz. The models are illuminated at incidence angles θ and ϕ (see Figure 2) both equaling 45° .

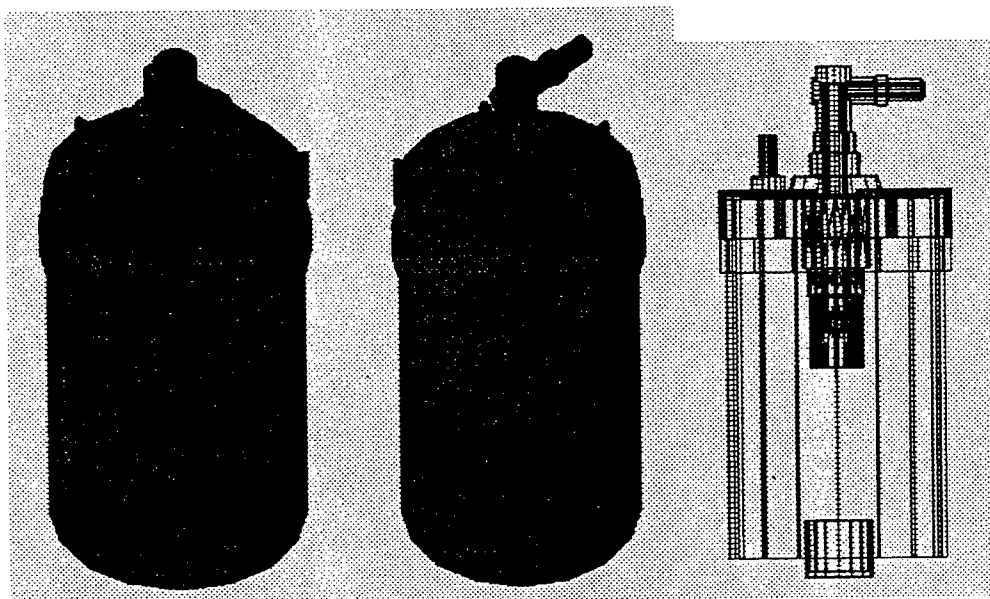


Figure 1. The Celestron-8 telescope.

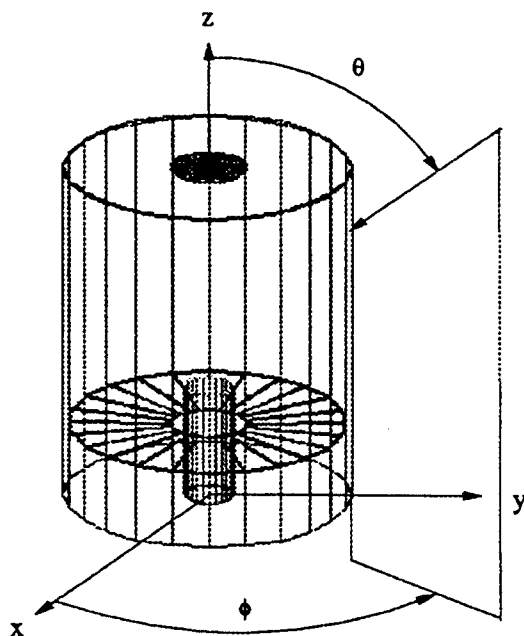


Figure 2. The CARLOS-3D model for the telescope.

Complementary to this analysis using CARLOS-3D, we were to use another software package called GEMACS (General Electromagnetic Model for the Analysis of Complex Systems) [3]. GEMACS uses three techniques which can be used to analyze a geometry, Geometrical Theory of Diffraction (GTD) and Method of Moments (MOM) to analyze exterior regions of a geometry and the Finite Difference method (FD) to analyze interior regions. We were interested in comparing the near field results obtained using GEMACS's MOM to the near field results obtained using MOM implemented in CARLOS-3D. Furthermore, the FD technique in GEMACS, was to be used to obtain cavity field results. Problems with both GEMACS and CARLOS-3D prevented a detailed study using GEMACS and the resulting comparison of the results from these codes.

Also of interest was the verification and validation of the results using different discretization techniques. We therefore discretized the model in Figure 2 for cell side lengths of $\lambda/5$ at 1.0 GHz. We analyzed this model for a frequency range of 0.5 GHz -3.0 GHz and compared the results for Quality Factor (Q), total energy, and exiting power for the $\lambda/5$ discretized model to those of $\lambda/10$ model. We noticed pronounced differences in the results over this frequency range. These results, other problems encountered while using CARLOS-3D, and methods for rectification of these problems are discussed in the following.

Dr. Buddy Coffey, creator of GEMACS, used the program to compare the field distribution within the open ended cylinder for $\lambda/20$ and $\lambda/40$ cell sides. The results proved to be identical. This led us to believe that GEMACS

should provide the most accurate results compared to CARLOS-3D. Even though GEMACS provided adequate results, our methodology of calculating total energy in the cavity results in rather long run times for GEMACS to generate the needed data. Dr. Coffey's suggestions for improvements which are outlined in Appendix A are provided for completeness.

Preliminary Observations

The model of Figure 2 was discretized into triangles having sides no longer than $\lambda/10$ in length at 1.0 GHz. The interior of this geometry was divided into cubes of side length 1.5 cm ($\lambda/10$ at 2.0 GHz). A total of 3717 grid points were created. These grid points are the locations of the field points calculated by CARLOS-3D. The model in Figure 2 was analyzed for a frequency range of 0.5 GHz up to 3.0 GHz. Using the fundamental equations describing the electric and magnetic energies along with power exiting the cavity, we calculated Q [4, 5] given by

$$Q = \omega \frac{U}{P} \quad (1)$$

where ω is the operating angular frequency, U is the total energy in the cavity and P is the power exiting the cavity. Figure 3 shows Q vs. frequency for the model of Figure 2. Resonances for various frequencies are evident and that the θ and ϕ polarizations follow a similar trend throughout the spectrum.

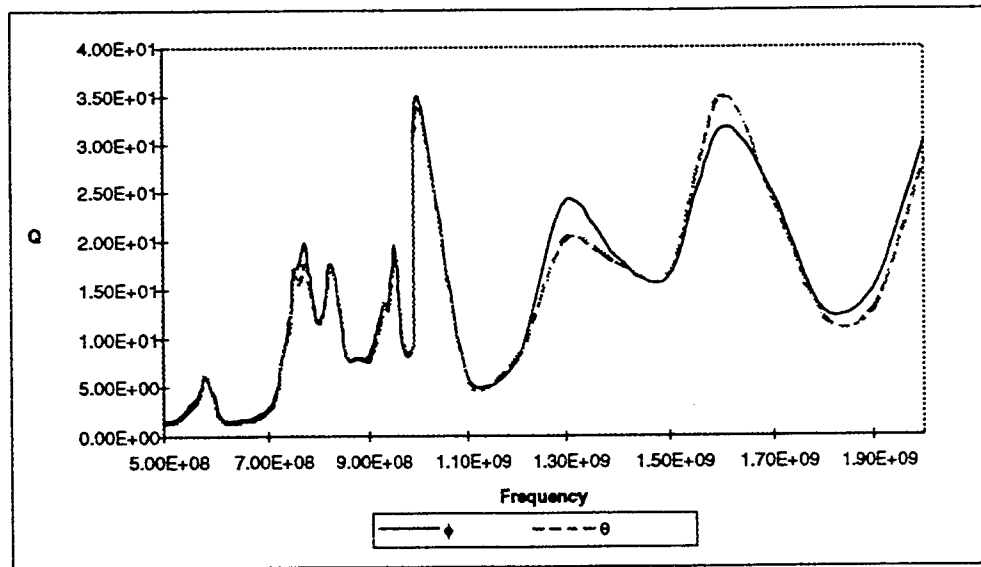


Figure 3. Quality factor vs. frequency for θ and ϕ polarization with $\lambda/10$ model

In order to investigate mesh sizes, we discretized the model of Figure 2 to $\lambda/5$ at 1.0 GHz and analyzed that model over frequency range of 0.5 GHz up to 3.0 GHz. Figure 4 shows that the results for the $\lambda/5$ and $\lambda/10$ models are profoundly different. These discrepancies led us to further investigate meshes using only the open cylinder model

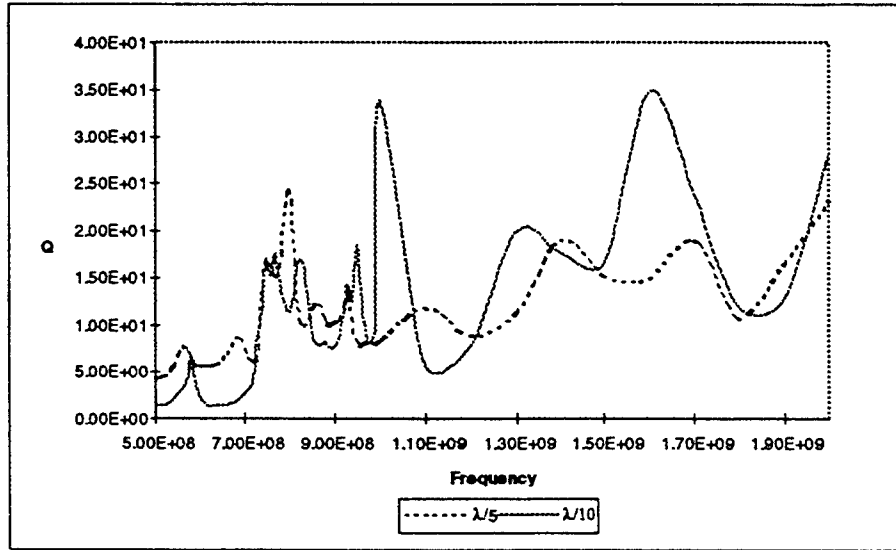


Figure 4. Telescope model quality factors, θ polarization, $\lambda/5$ and $\lambda/10$ models.

with two different discretization schemes, $\lambda/10$ and $\lambda/20$. We analyzed these models for a frequency range of 0.1 GHz to 2.0 GHz. Figure 5 shows these results again with large differences.

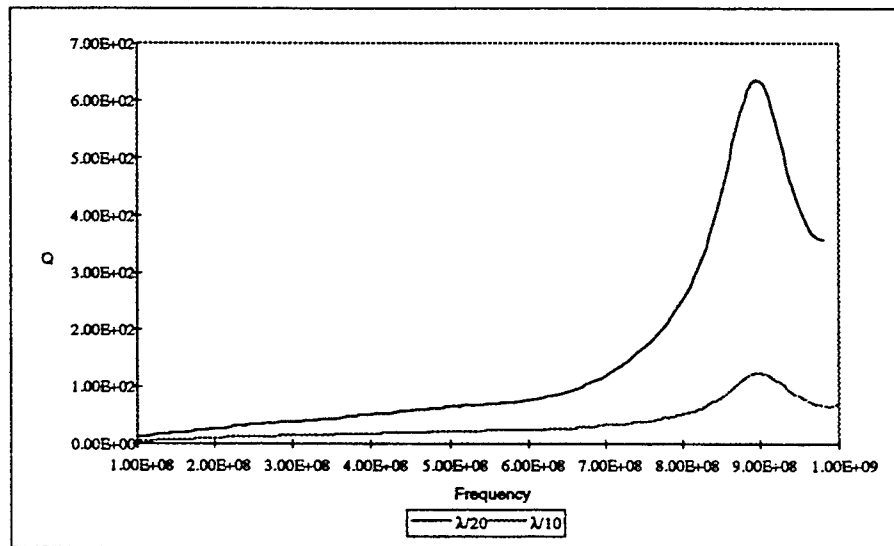


Figure 5. Open cylinder model quality factors, θ polarization, $\lambda/10$ and $\lambda/20$ models.

At this point it was clear that we had to investigate fully the problems with CARLOS-3D that lead to different results for different meshes. This final report addresses five areas of study: (1) the field calculations in CARLOS-3D, (2) non-uniform meshing schemes, (3) the matrix solution in CARLOS-3D and how it might be improved using iterative refinement and improved Gaussian Quadrature, (4) High Frequency Structure Simulator (HFSS) models and results, and (5) testing results.

Field Calculation Method and Results

We first approached the meshing problem by comparing the field quantities for the open cylinder models ($\lambda/10$ and $\lambda/18$). Doing this, we realized that there were large differences in field values near the cylinder walls compared to those field values more towards the center of the cylinder. During conversations with the creators of CARLOS-3D, we discovered that even though the currents throughout each cell are known, CARLOS-3D calculates near fields using values of currents at the centroid of each triangular cell. In this scenario, each current source will now act similar to an infinitesimal dipole. Therefore, the field quantities are calculated using [6]

$$E_z(x, y, z) \propto \sum_{n=1}^N \frac{e^{-jk r_n}}{k r_n^3} \cos(\theta) J_n \quad (2)$$

where k is the propagation constant, J_n is the n_{th} surface current, and r_n is the distance from the n_{th} surface current to the field point. This model differs significantly from one that can be used with takes advantage of the current distributions that are known. In that case the distance r from the field point varies according to where the current is calculated on the surface. The distance r is a very critical quantity since for small r , the inverse cubic variation will have a large effect on the field calculations. Figures 6, 7 and 8 show the percent difference between the $\lambda/10$ and $\lambda/18$ models in the electric field quantities along the x-axis at the end cap, the center, and at the aperture.

Iterative refinement (see succeeding sections) was implemented for creating the figures. It is evident in Figures 6, 7 and 8, that the biggest differences are close to the cavity boundary. This error is caused by the crude approximation used to calculate the near fields.

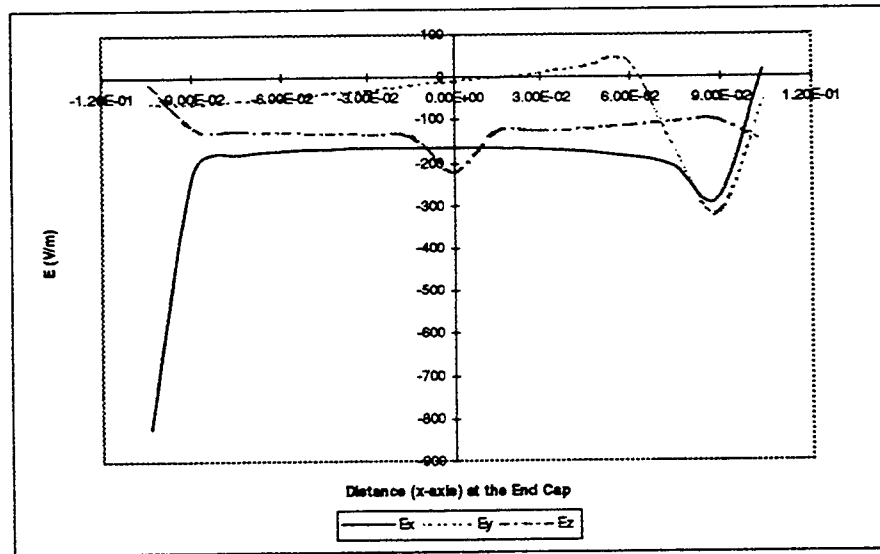


Figure 6. Percent difference in the electric field quantities for both models, end cap.

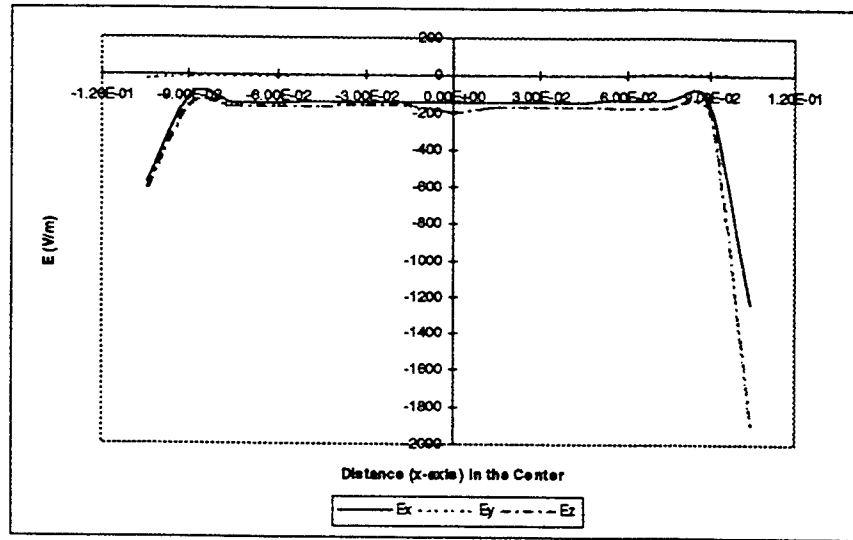


Figure 7. Percent difference in the electric field quantities for both models, center.

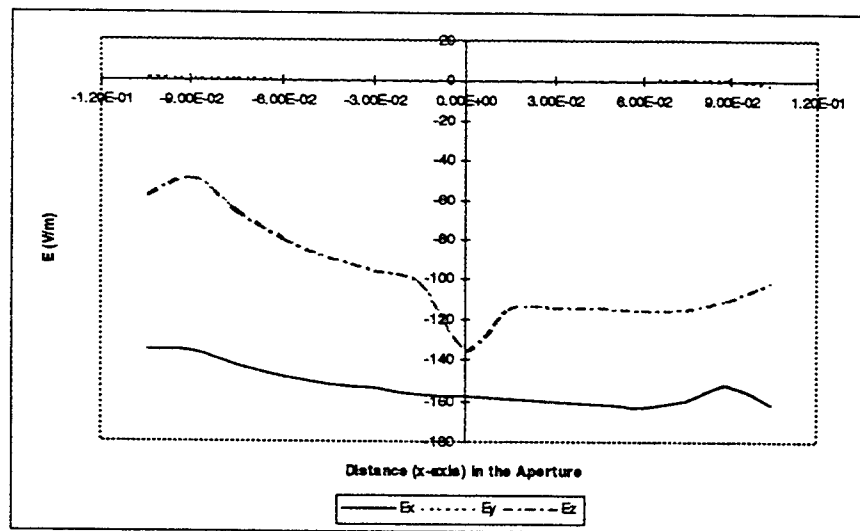


Figure 8. Percent difference in the electric field quantities for both models, at the aperture.

A more suitable approach would be to use a solution technique which considers the changes in r along with changes in the current density. This solution is given by [7]

$$\begin{aligned} \mathbf{E}^*(\mathbf{r}) = & -j\omega\mu \int_{\partial V_s} \mathbf{J}_s(\mathbf{r}') g(\mathbf{r}, \mathbf{r}') ds' \\ & + \nabla \int_{\partial V_s} \frac{1}{j\omega\epsilon} \nabla' \cdot \mathbf{J}_s(\mathbf{r}') g(\mathbf{r}, \mathbf{r}') ds' \end{aligned} \quad (3)$$

where

$$g(r, r') = \frac{e^{-jk|r-r'|}}{4\pi|r-r'|} \quad (4)$$

and ω is the angular operating frequency, ϵ and μ are the permittivity and permeability of free space, respectively. The functional dependence of $J_s(r)$ is determined by the basis functions (for CARLOS-3D these are linear) and the functional dependence of $g(r, r')$ is given by equation (4).

Meshing Methods And Results

In the neighborhood of a conducting edge (Figure 9), electric and magnetic field components behave in a particular manner dependent on the wedge angle ϕ_0 .

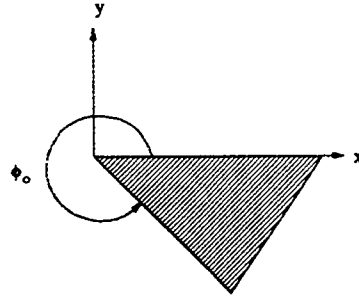


Figure 9. Wedge angle ϕ_0 .

From [8] the behavior of the electric field and the surface current density in the vicinity of the edge is

$$E \propto \rho^\nu \sin(\nu\phi) \quad J_z \propto \rho^{\nu-1} \quad (5)$$

where

$$\nu = \frac{\pi}{\phi_0} \quad (6)$$

At the top edge of the open cylinder, $\phi_0=2\pi$, and at the bottom edge (at the end cap) $\phi_0=3\pi/2$. For our geometry then, at the top edge, $J_\phi \propto \rho^{-1/2}$ and at the bottom edge, $J_\phi \propto \rho^{-2/3}$. The edge condition is an additional constraint that is needed for a unique solution whenever a geometric singularity exists and states that the energy density in the vicinity of any geometrical singularity must be integrable. That is

$$\text{Re} \int_S \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{S} \rightarrow 0 \quad (7)$$

as the surface S enclosing the edge shrinks to the edge. This condition dictates that the edge shall not radiate any energy because it is not a source.

In order to accurately model the edge surface currents we used a meshing scheme created in Mathematica as shown in Figure 10. Figure 11 shows the z , ρ and ϕ directed surface currents for the meshing scheme of Figure 10.

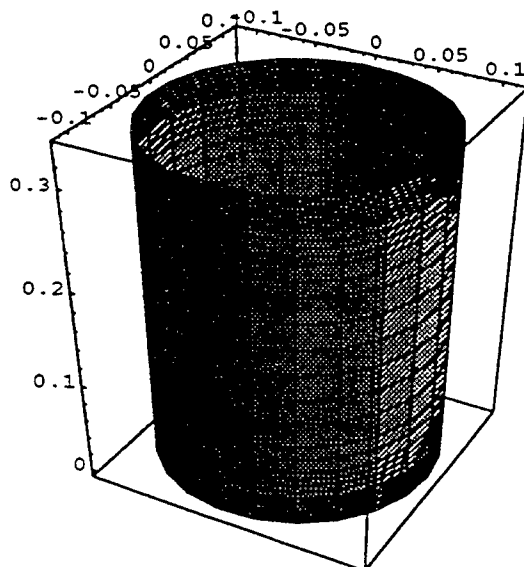


Figure 10. Non-uniform model, using (8,4,2,6,2,4,8) discretization technique.

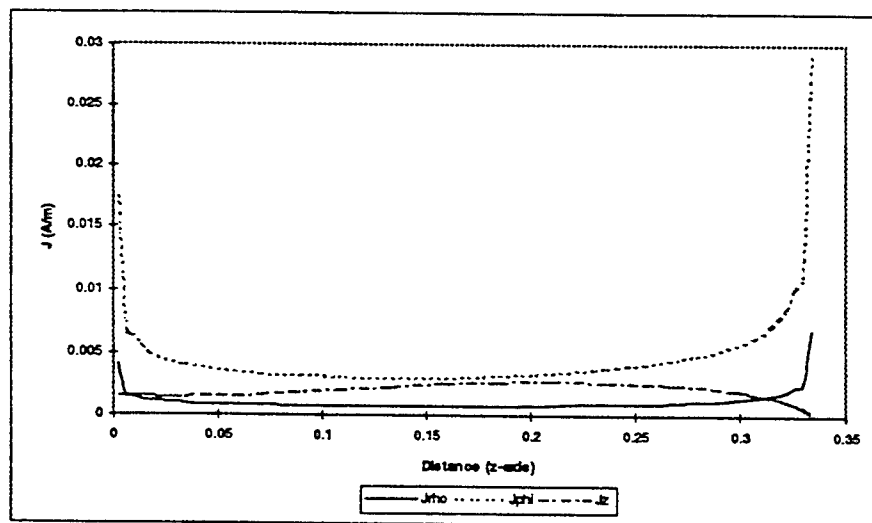


Figure 11. Surface currents for the meshing scheme of Figure 10. ϕ polarization, $\phi = 5^\circ$, $f = 500$ MHz. The incidence angle is $\theta = 45^\circ$, $\phi = 0^\circ$

In addition to the accurate modeling of edge currents, we also looked at the problem of generating a mesh for a circular disk. For problems of circular symmetry, the cells near the center of circles become exceedingly narrow. A simple transition to a non-symmetric grid, as shown in Figure 12, would be desirable to maintain reasonable cell sizes. Dr. Pat McDaniel wrote a program to accomplish this task and also give a finer mesh for cells closer to the edges. The program provided meshes for the sides of the cylinder in a more efficient way than shown in Figure 10.

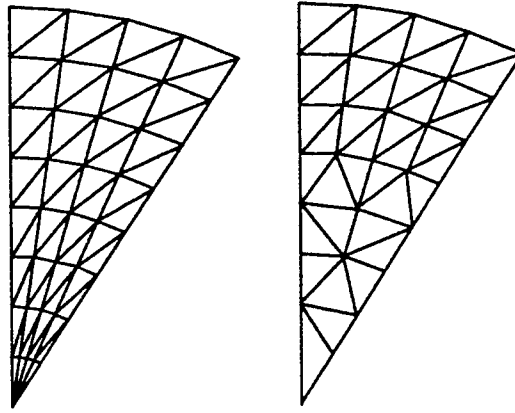


Figure 12. Alternate gridding scheme for circular geometry.

The basic concept for determining the width of the edges of the cells for accurate edge current modeling is to maintain a constant current through each cell edge as you move further and further away from the edge. In this way the cell widths can be determined and a minimum amount of cells will be generated. Figure 13 shows some results for this meshing scheme. The meshing scheme of Figure 10 (without mesh refinement near the edge of the end cap) generated 2507 unknowns in CARLOS-3D. This is compared to Dr. McDaniel's meshing which generated 2034 unknowns in CARLOS-3D. Both are for the $\lambda/18$ model.

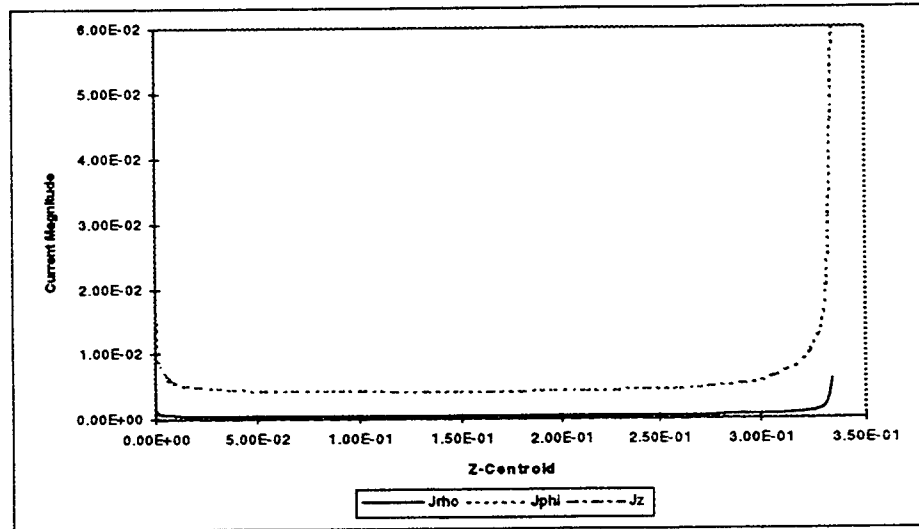


Figure 13. Surface currents using Dr. McDaniel's meshing scheme: $\lambda/18$ model, incidence angle is $\theta = 90^\circ$, $\phi = 0^\circ$, ϕ polarization, $f = 500$ MHz.

At this point, we decided to do another frequency sweep with Dr. McDaniel's meshes. The results for Q , power exiting the cavity, and energy for the θ polarization (results are similar for the ϕ polarization) are given in Figures 14 - 16.

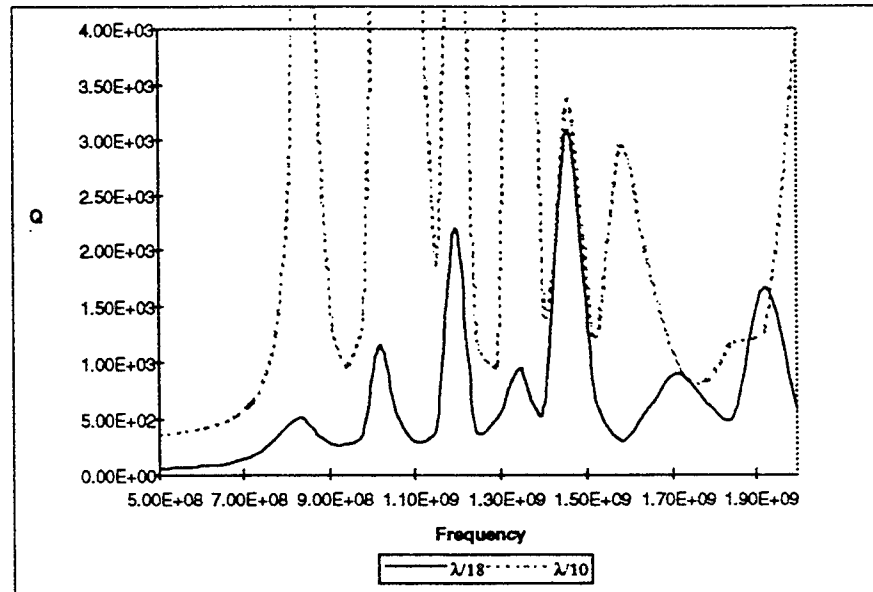


Figure 14. Q for the non-uniform meshes, θ and ϕ polarizations.

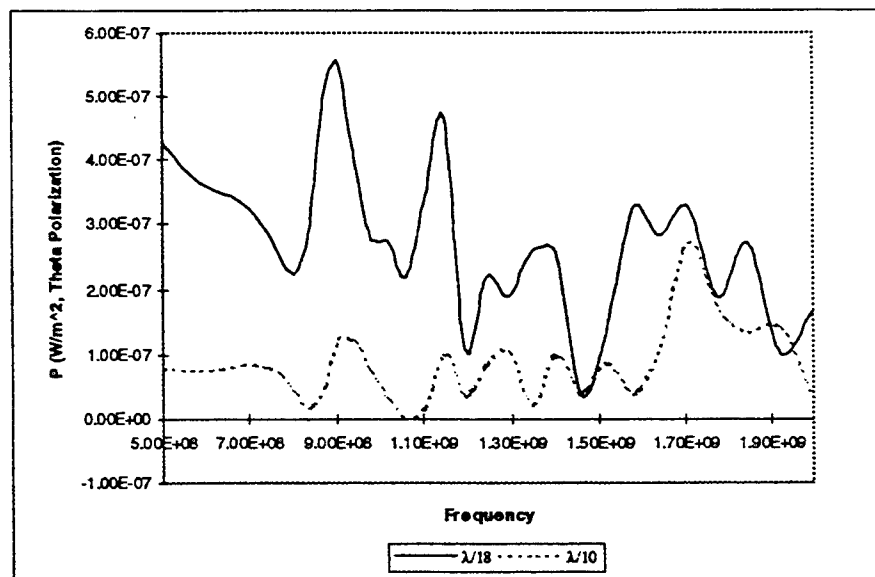


Figure 15. Power for the non-uniform meshes.

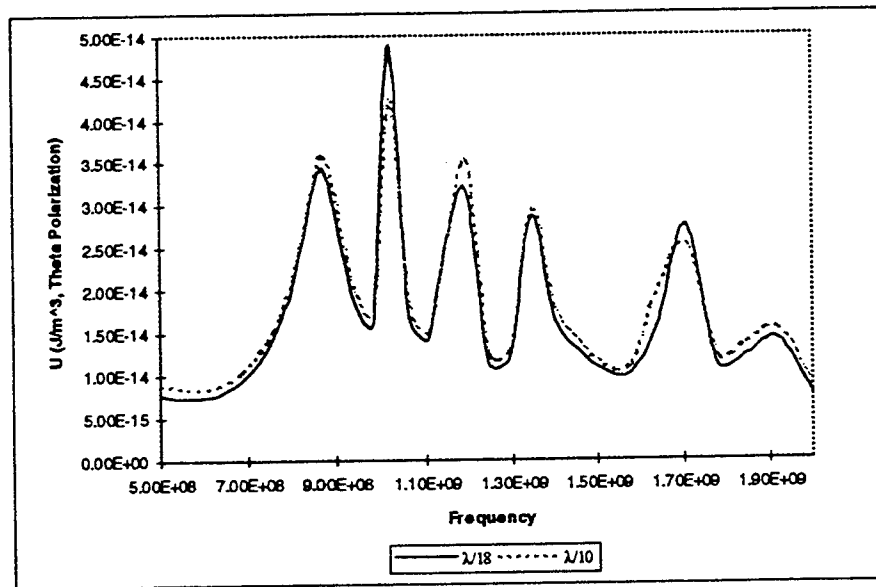


Figure 16. Energy for the non-uniform meshes.

The energy calculations for the two meshes are much better, but the differences in the power calculations are pronounced. We can see now that the major contributors to different Qs for different meshes are the power differences in the aperture. Graphs of the differences of the power calculations over the aperture for the case of Figures 14 - 16 and for the ϕ polarization are shown in Figure 17. These differences can be explained by the fact that our power calculations are at 177 points in the aperture and are vector calculations giving positive and negative results which are summed over the area of the aperture. This process is inherently prone to more error than the energy calculations which are calculations of field magnitudes (all positive values) over 3717 points that represent the volume of the cavity

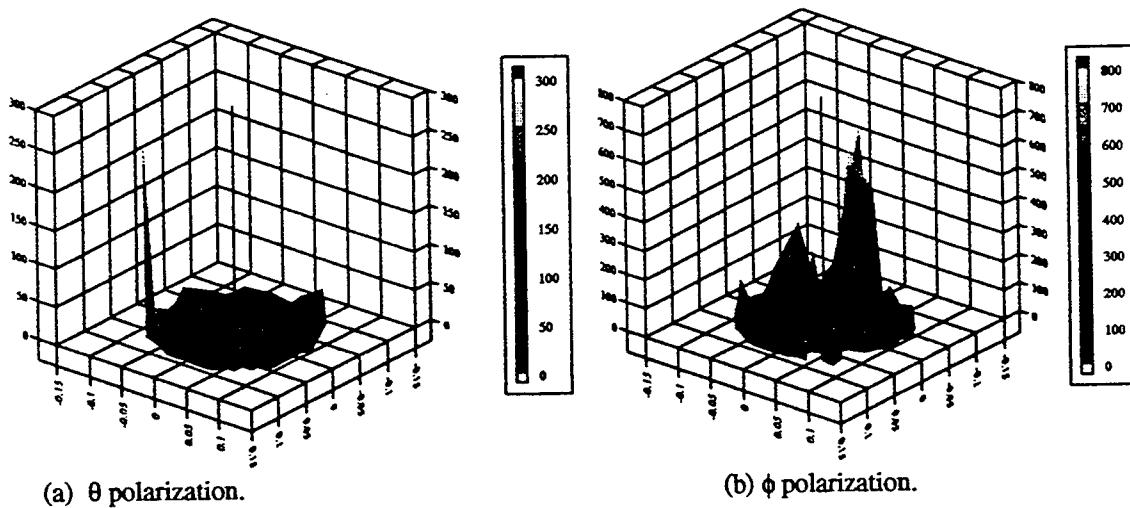


Figure 17. % difference in aperture power magnitude for the non-uniform meshes.

Accuracy and results for high condition number

Other numerical difficulties resulting from using CARLOS-3D, may be due to the large condition number of the impedance matrix for large models. The condition number of a matrix may be defined as its largest element value divided by its smallest [9]. A matrix is typically considered ill-conditioned if its corresponding inverse matrix is numerically unstable. Taking another point of view, a zero, or near zero determinant indicates that the matrix has, in effect, two rows that are identical or approximately identical. In a system of linear algebraic equations, a zero determinant (two identical rows) means that two of these equations are either redundant or inconsistent. In an approximate-solution method such as MOM, the redundancy and inconsistency in a system of linear equations manifest themselves as numerical instabilities. A large condition number indicates that the matrix may be ill-conditioned. The condition numbers are not always reliable in predicting the stability of a matrix. We have to know how large the condition number must be before we declare the matrix ill-conditioned. A practical rule of thumb is that the matrix is ill-conditioned if the inverse of the condition number approaches the computer's floating-point precision. For an ill-conditioned matrix equation, iterative methods are generally more suitable. An iterative refinement technique was implemented by Dr. Pat McDaniel and is discussed here.

It is not always easy to obtain greater precision for the solution of a linear set than the precision of your computer's floating-point numbers. In direct methods of solution, round-off errors accumulate, and they are magnified to the extent that the matrix is close to singular. Under these conditions, matrices, which were far from singular, can lose two or three significant figures. Iterative refinement (improvement) is an approach to restore full machine precision [10]. Suppose that a vector \mathbf{x} is the exact solution of a linear set

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} \quad (8)$$

however, the exact solution \mathbf{x} is not known. What is known, in this case, is a slightly wrong solution $\mathbf{x} + \delta\mathbf{x}$, where $\delta\mathbf{x}$ is the unknown error. Therefore, from equation (8), it is evident that when multiplied by the matrix \mathbf{A} , the slightly wrong solution gives a product slightly discrepant from the desired right-hand side \mathbf{b} , such that

$$\mathbf{A} \cdot (\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b} \quad (9)$$

Now, Subtracting equation (8) from (9) gives

$$\mathbf{A} \cdot \delta\mathbf{x} = \delta\mathbf{b} \quad (10)$$

Using equation (10) to solve for $\delta\mathbf{b}$ and then substituting that into equation (9) gives

$$\mathbf{A} \cdot \delta\mathbf{x} = \mathbf{A} \cdot (\mathbf{x} + \delta\mathbf{x}) - \mathbf{b} \quad (11)$$

The right side of equation (11) is known, since $\mathbf{x} + \delta\mathbf{x}$ is the wrong solution that requires refinement. Therefore solving equation (11) for $\delta\mathbf{x}$, subtracting it from the wrong solution will provide an improved solution.

Numerical Integration Of The Matrix elements

The integrals of a function are approximated by the sum of its functional values at a set of equally spaced points, multiplied by certain aptly chosen weighting coefficients. The idea of Gaussian Quadrature is to give the user the freedom to choose not only the weighting coefficients, but also the location of the abscissas at which the function is to be evaluated, hence they will no longer be equally spaced. Using Gaussian Quadrature formulas, one can arrange the choice of weights and abscissas to make the integral exact for a class of integrands. Therefore, the weighting function $W(x)$ can be chosen to remove integrable singularities from the desired integral. Hence, one can find a set of weights w_i and abscissas x_i such that the approximation

$$\int_a^b W(x)f(x)dx \approx \sum_{i=1}^N w_i f(x_i) \quad (12)$$

is exact if $f(x)$ is a polynomial. CARLOS-3D is a MOM code and hence its primary calculations are surface currents. The option exists in CARLOS-3D to change namelist variables to improve the Gaussian Quadrature integrations. Figures 18 - 23 present the percent differences in currents for iterative refinement versus iterative refinement and improved Gaussian Quadrature. Note that the differences are greater for currents near the rectangular edge at the end cap.

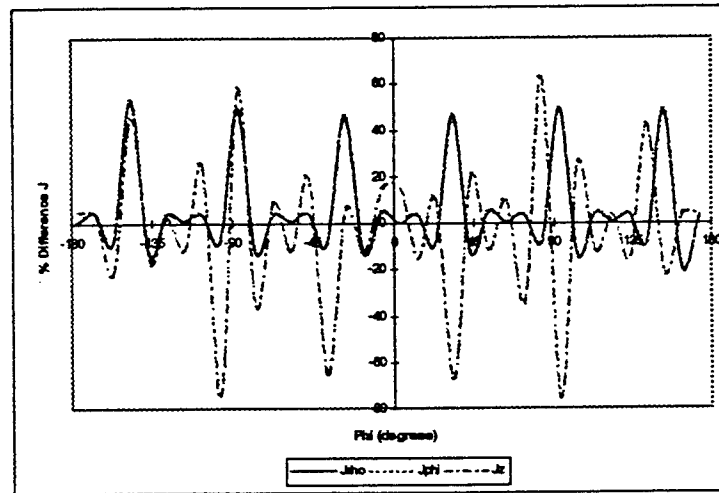


Figure 18. Percent difference in currents for the cylinder side at the end cap for the $\lambda/18$ mesh, $f = 500$ MHz, $\theta = 90^\circ$ and $\phi = 0^\circ$.

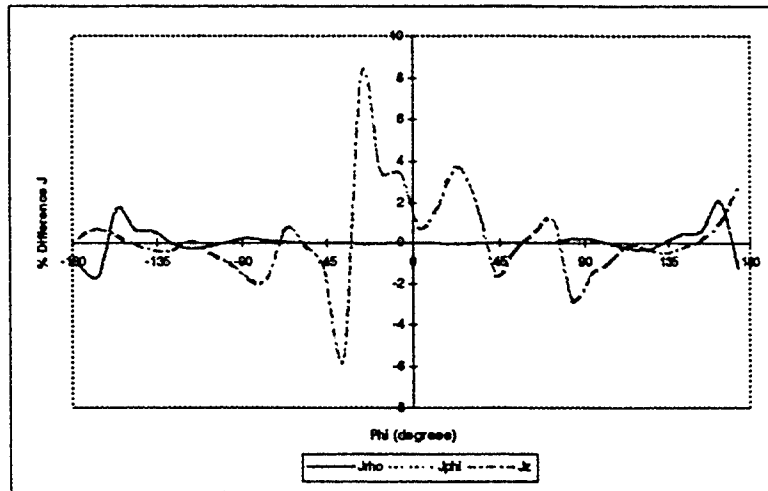


Figure 19. Percent difference in currents for the cylinder side at the center for the $\lambda/18$ mesh, $f = 500$ MHz, $\theta = 90^\circ$ and $\phi = 0^\circ$.

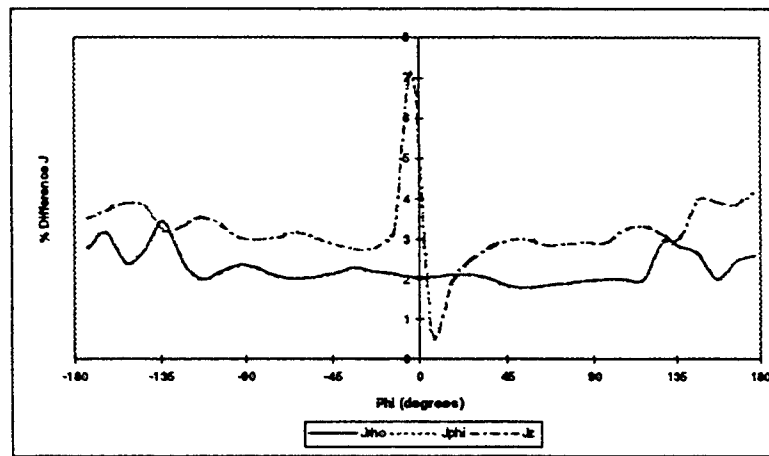


Figure 20. Percent difference in currents for the cylinder side at the aperture for the $\lambda/18$ mesh, $f = 500$ MHz, $\theta = 90^\circ$ and $\phi = 0^\circ$.

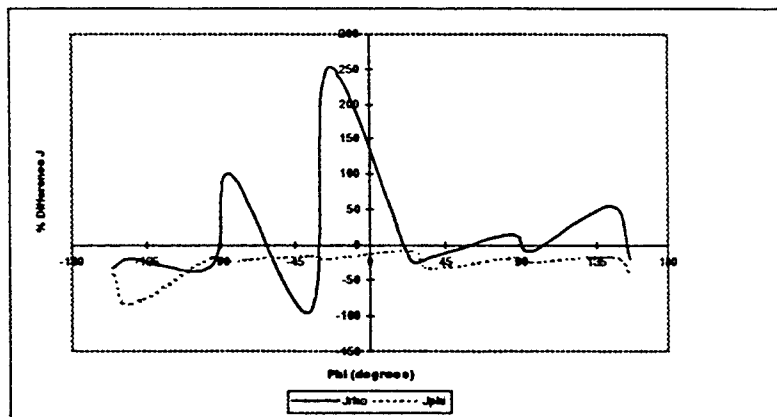


Figure 21. Percent difference in currents for the cylinder end cap near the edge for the $\lambda/18$ mesh, $f = 500$ MHz, $\theta = 90^\circ$ and $\phi = 0^\circ$.

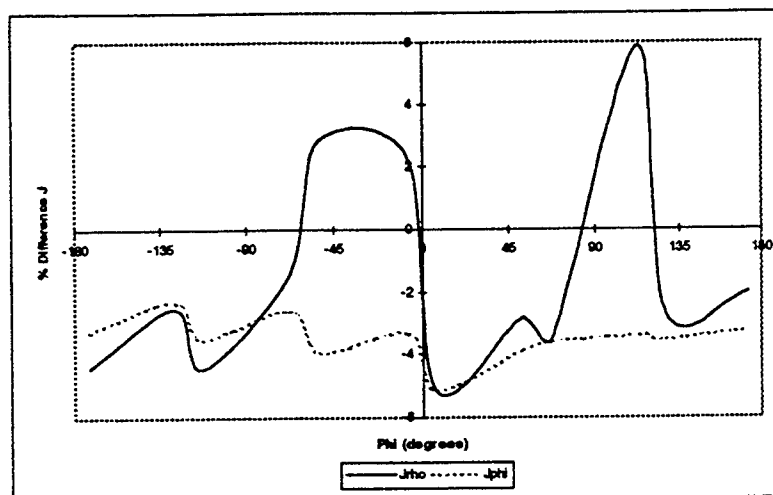


Figure 22. Percent difference in currents for the cylinder end cap at the center for the $\lambda/18$ mesh, $f = 500$ MHz, $\theta = 90^\circ$ and $\phi = 0^\circ$.

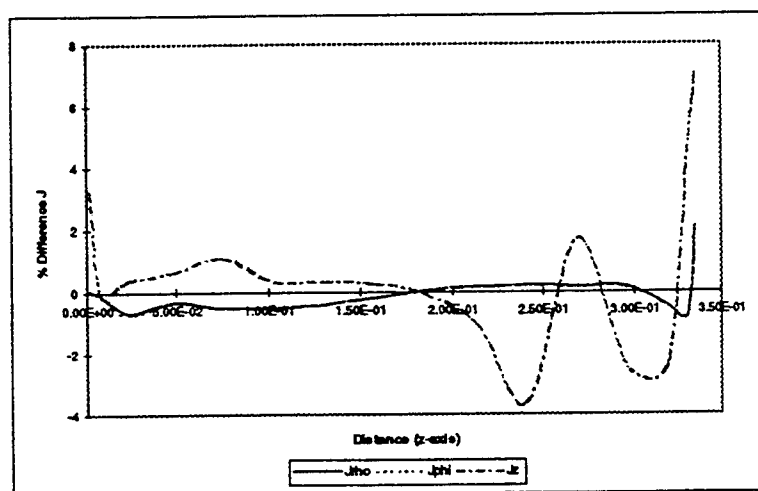


Figure 23. Percent difference in currents for the cylinder side along the z axis at $\phi = 3^\circ$ for the $\lambda/18$ mesh, $f = 500$ MHz, $\theta = 90^\circ$ and $\phi = 0^\circ$

HFSS description and results

The Hewlett Packard (HP) High Frequency Structure Simulator (HFSS) is a three dimensional finite element simulator. In the simulator, the geometric model is automatically divided into a large number of tetrahedra which is called the finite element mesh. Within each tetrahedron, the vector field components perpendicular to the edges and tangential to the edges are mathematically modeled using polynomials. The finer the mesh, the more accurate the solution. Of course, the finer the mesh, the greater the demands on computational resources in terms of memory, disk space and solution time. In order to find a good solution to the problem and minimize computer resources, HFSS uses "an iterative process in which the mesh is automatically refined in critical regions." The simulator first solves the problem using a course initial mesh, refines the mesh in areas of high error density,

generates a new solution, etc., until selected S-parameters converge to within a desired limit, or until all computer resources have been exhausted.

The computer that we use to run HFSS is an IBM RS6000 model 570 PowerServer with 128 MBytes of RAM and a 2 GByte hard drive. We have access to all of the RAM, about 200 MBytes of swap space, and 340 MBytes of disk space for user files. During the solution process, a convergence menu can be monitored containing memory and file size values. Two matrices are written to disk (user files); the A matrix is the “full matrix” and the U matrix is the “upper triangular matrix”. All other memory size values indicated on the convergence menu are RAM usage. As stated above, the iterative process of mesh refinement is continued until convergence, or until the computer resources are exhausted. If the disk space is exhausted (which was the limiting resource for our models), the solution did not converge.

A total of eight projects were created for HFSS. They were:

1. cyltest.pjt - A cylindrical waveguide port excited cylinder which is the same size as the outer cylinder of the Celestron Telescope (see Figure 24.) The radiation boundary was positioned 150 mm from the cylinder boundary. This was done so that the radiation boundary would be $1/4 \lambda$ (at the lowest frequency of interest - 0.5 GHz) from the radiating surface as suggested by HP.

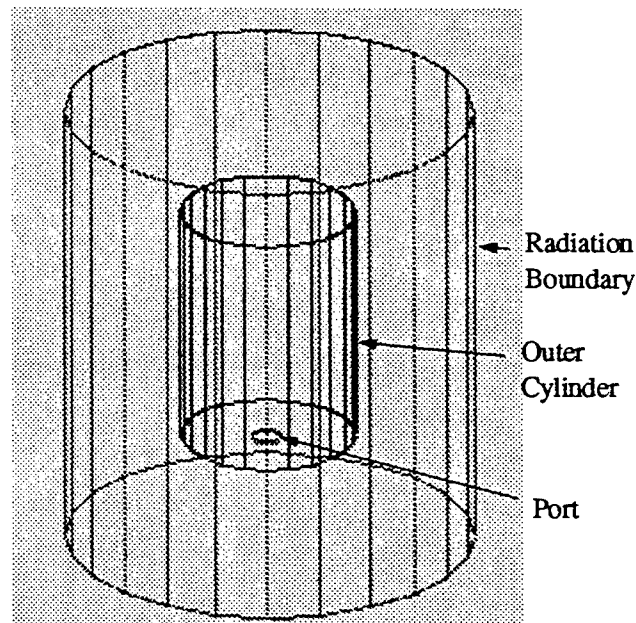


Figure 24. The cylindrical waveguide port excited cylinder.

2. cylinder.pjt - A coaxial probe excited cylinder which is the same as shown in Figure 24 with the exception that the port is replaced by a coaxial probe which has its center conductor inserted along the z axis into the cylinder a distance of 3 cm. The probe is 1 mm in radius. The outer conductor of the coaxial port is 2.3 mm in radius.
3. telescop.pjt - A full model of the telescope with the exception that the large mirror is flat. We could not get the 2-D cross-section to close using the true radius of curvature. Since the 2D cross-section would not close, the 3D mirror could not be created. Thus, a flat mirror with approximate tilt was created. The center of the mirror is at $z = 72$ mm and the outside is at $z = 73.8$ mm. This model is excited by a cylindrical waveguide port located at the eye piece.
4. teletest.pjt - A full model of the telescope (Figure 25) as detailed above in telescop.pjt with the exception that the eye piece is open and a coaxial probe and port as described in cylinder.pjt is located at $z = 167.5$ mm.
5. telehf.pjt - A model of the telescope excluding the volume behind the large mirror and with the radiation boundary located 17 mm from the large cylinder boundary. This model was specifically created at a frequency of 4.5 GHz in an attempt to have a model at a frequency higher than the cutoff frequency of the inner cylinder (4.4 GHz).
6. cyltest2.pjt - An open cylinder excited by a square coaxial probe at the center of the cylindrical shell with the probe oriented along the x axis. The square cross-section probe was used to alleviate the problem of a high mesh density at the small cylindrical probe. The inner conductor (probe) of the coaxial feed is 4 mm wide, the outer conductor has a 9.2 mm inside width. The radiation boundary was chosen to accommodate a 0.75 GHz - 1.5 GHz frequency range.
7. cyltest3.pjt - This project is the same as cyltest2.pjt with the exception that the radiation boundary is closer to accommodate a 1.5 GHz - 2 GHz frequency range.

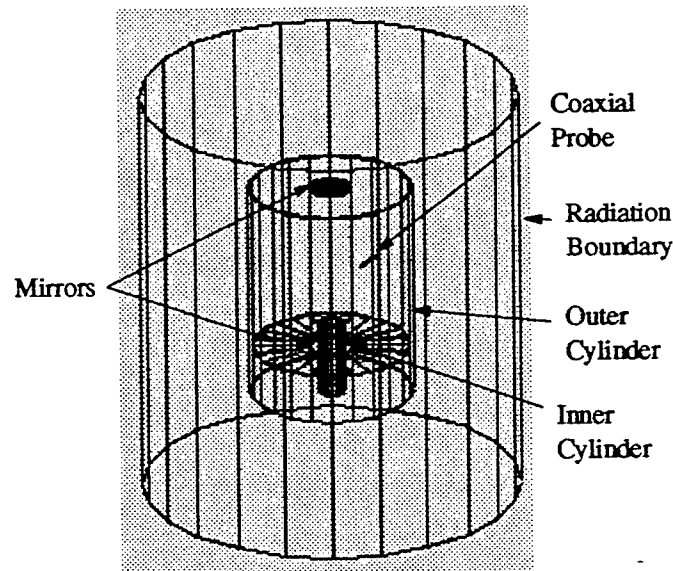


Figure 25. The coaxial probe excited telescope.

8. telerad.pjt - This model was like that for telescop.pjt except that the radiation boundary condition was enforced at the aperture of the telescope. This in effect eliminates any fields due to currents on the outside of the telescope. This has some effect on the fields inside the telescope but should provide minimal error. The goal was to get a model that would work at a frequency above the cutoff frequency of the small inner cylinder. HFSS failed to refine the mesh on the second pass. We think that this is an indication that we ran out of RAM (>128 MBytes.)

Once again, we ran up against memory (RAM) and disk storage limitations. Table 1 shows the details of the various projects and models at the given frequencies along with the memory requirements (disk space and RAM) of each. We were unable to converge to a 2 GHz model for the telescop.pjt or to even a 1 GHz model for the teletest.pjt. An attempt at a 870 MHz model was made because a resonance at that frequency has been measured. The model did not converge. Also, the telehf.pjt model did not converge.

Table 1. Summary of HFSS projects and resource requirements.

Name of Project	Frequency	A matrix Disk Space (K)	U matrix Disk Space (K)	csolve RAM (K)	Total Time
cyltest.pjt	1 GHz	16,302	65,965	37,340	31:39
	2 GHz	25,717	134,732	67,996	1:25:54
cylinder.pjt	500 MHz	23,588	123,220	65,820	1:22:42
telescop.pjt	1 GHz	14,130	77,376	66,332	17:13
	1.5 GHz	23,122	164,935	103,132	41:48
	2 GHz	---	>340,000	---	---
teletest.pjt	1 GHz	---	>340,000	---	---
	870 MHz	---	>340,000	---	---
telehf.pjt	4.5 GHz	---	>340,000	---	---
cyltest2.pjt	1.5 GHz	34,083	211,128	107,996	2:40:00
cyltest3.pjt	2 GHz	39,834	243,013	114,396	3:03:00
telerad.pjt	4.5 GHz	---	---	---	---

Figures 26 - 29 show results for telescop.pjt at 1.18 GHz. Even though a good model was not generated at 4.5 GHz for this project, we looked at the radiation results (Figure 30) for the best model generated. When compared to the radiation plot of Figure 29, it is clear that once we get above the cutoff frequency of the inner cylinder, the radiation increases substantially. It is interesting to note that even though the frequency of excitation (1.18 GHz) is below the cutoff frequency of the inner cylinder waveguide, there are calculable fields at the end of the inner cylinder and at the aperture of the large cylinder (see Figure 26.) The magnetic field intensity magnitude plots of Figures 27 and 28 also show significant fields behind the large mirror.



Figure 26. Electric field - cylinder openings.

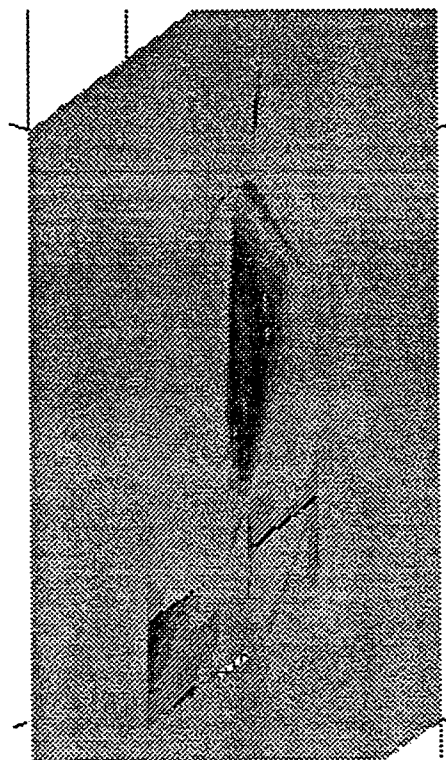


Figure 27. Magnetic field - x axis.

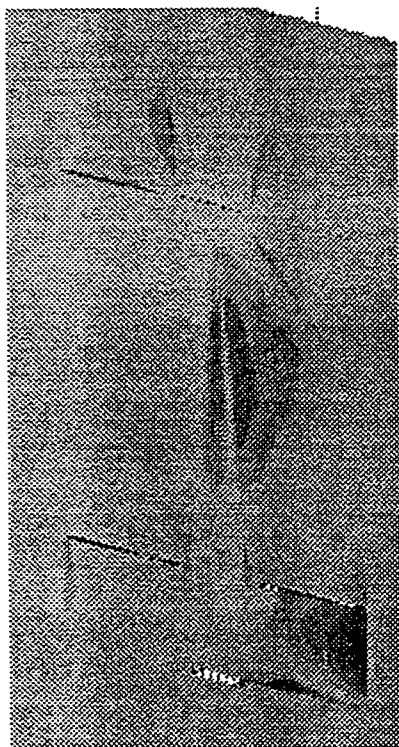


Figure 28. Magnetic field - y axis.

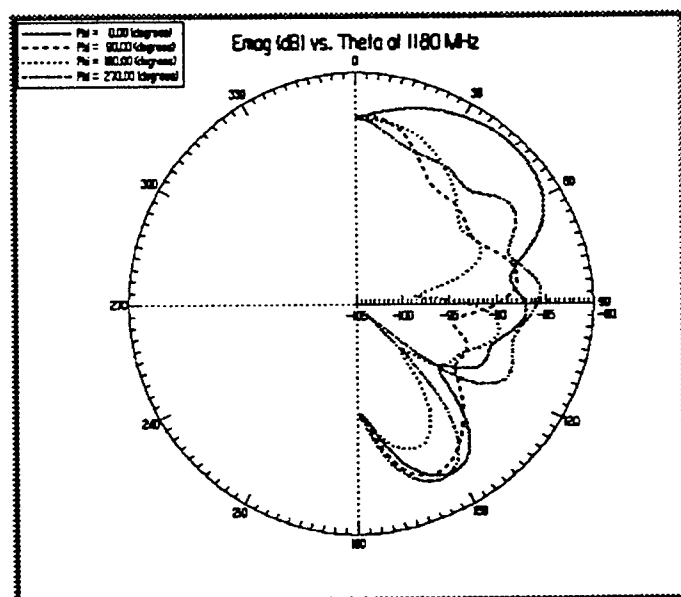


Figure 29. Radiation at 1.18 Ghz.

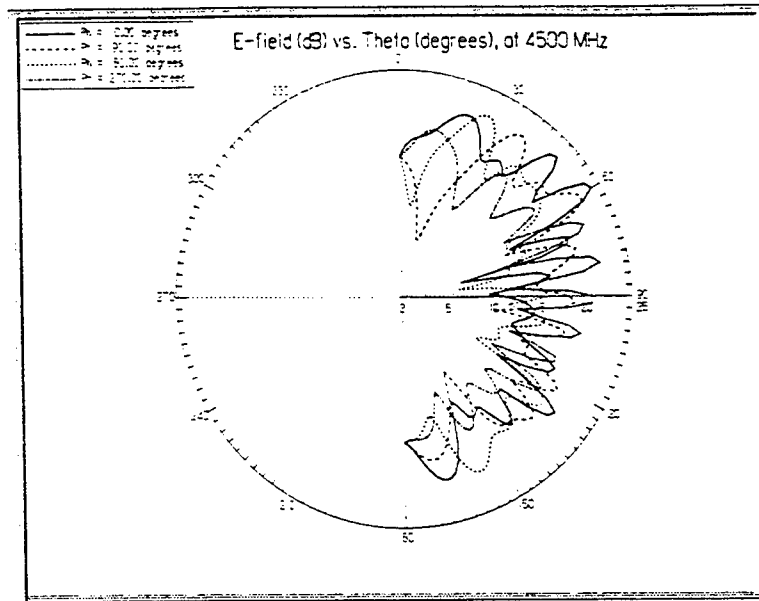


Figure 30. Radiation at 4.5 Ghz.

HFSS Results vs. Physical Testing

The most successful models were cyltest2.pjt and cyltest3.pjt. Results for the fields and radiation plots at the two major resonances are shown in Figures 31 - 36. Figures 37 - 39 show results for S_{11} from HFSS and two tests of the telescope with a coaxial probe positioned as for cyltest2.pjt. For the test results, a 3 cm probe was placed at the center of the outer shell and a set of S_{11} measurements were taken. These measurements were done once having a conducting plate at the aperture of the telescope and then same measurements were done having no cover plate at the aperture. In both cases a receiver horn antenna was placed at the eyepiece of the telescope. The antenna was 0.88 inches in full length with aperture dimensions of 0.49 inches in length and 0.40 inches in height.

As a reference, the TE_{211} and TE_{311} modes for a closed cylindrical cavity of the dimensions of our open cavity resonate at 1.36 GHz and 1.822 GHz. The electric field plots at 1.35 GHz and 1.8 GHz clearly indicate that the transverse modes are TE_{21} and TE_{31} respectively. Figure 39 shows results for S_{11} from a test of the telescope with a conducting cover plate at the aperture. These results when compared to the results of Figure 38 show that the resonances are at the same location but of different depths. So our simple model of an open cylinder accurately models the telescope for these low frequency resonances.

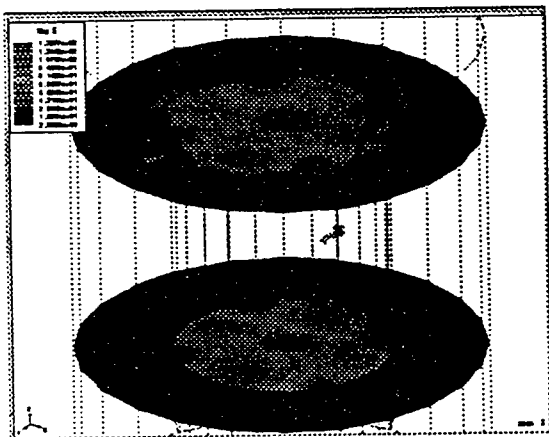


Figure 31 Electric field at 1.35 GHz.

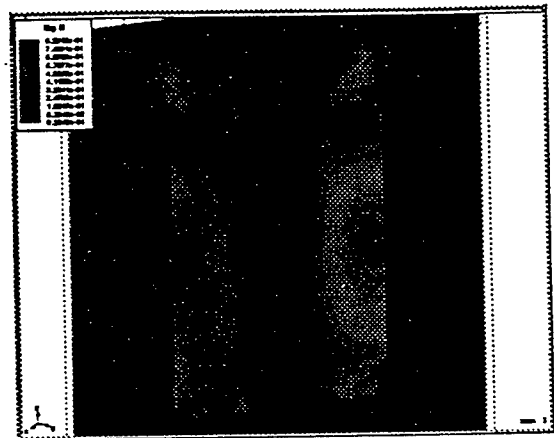


Figure 32. Magnetic field at 1.35 GHz, $\phi=45^\circ$.

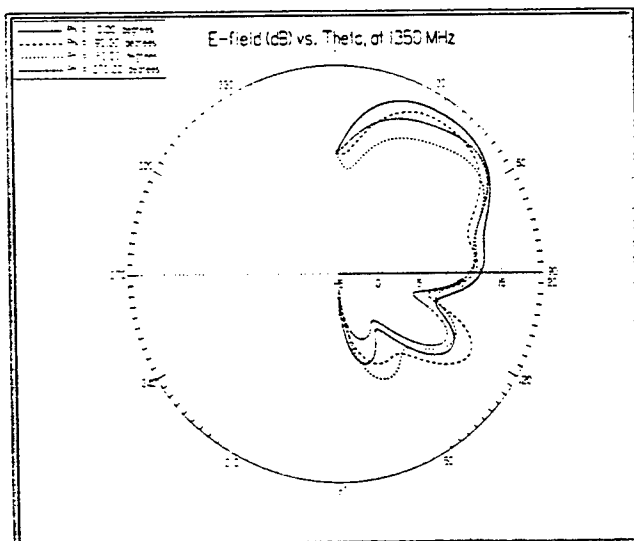


Figure 33. Radiation at 1.35 GHz.

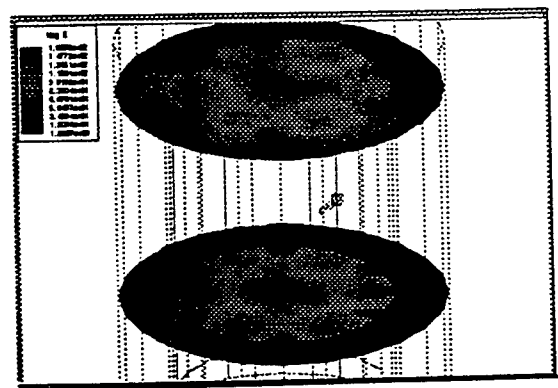


Figure 34. Electric field at 1.80 GHz.

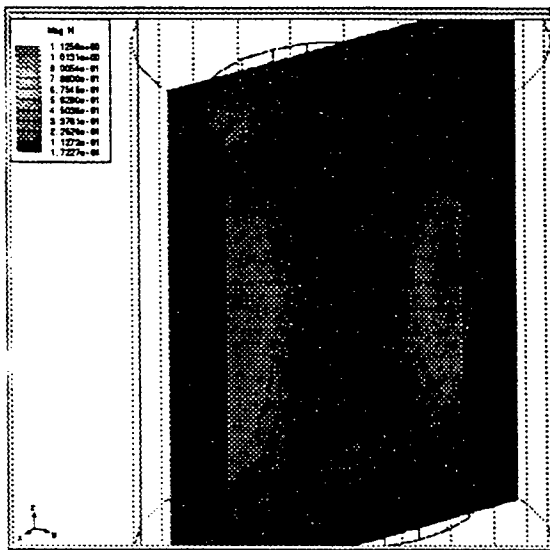


Figure 35. Magnetic field at 1.80 GHz, $\phi = 30^\circ$ plane.

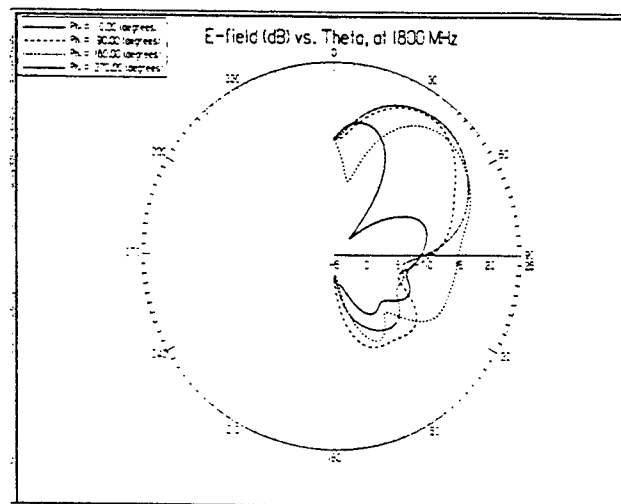


Figure 36. Radiation at 1.80 GHz.

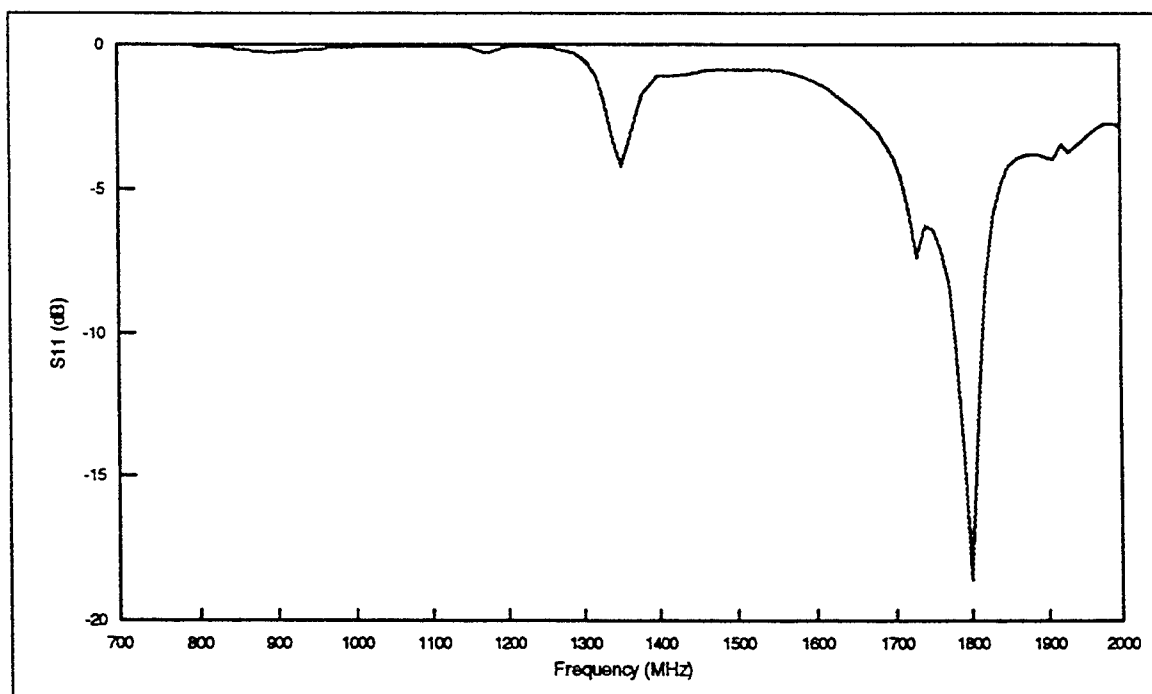


Figure 37. HFSS results for the probe excited open cylinder (10 MHz resolution.)

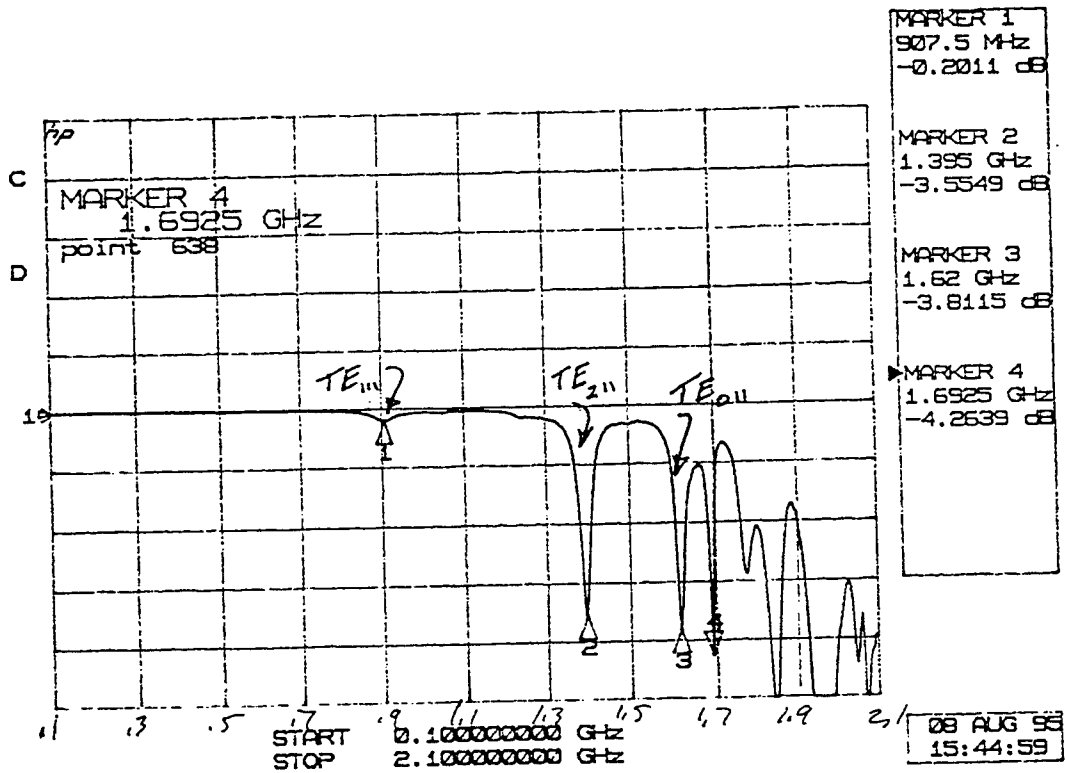


Figure 38. Measured S_{11} results for the open telescope.

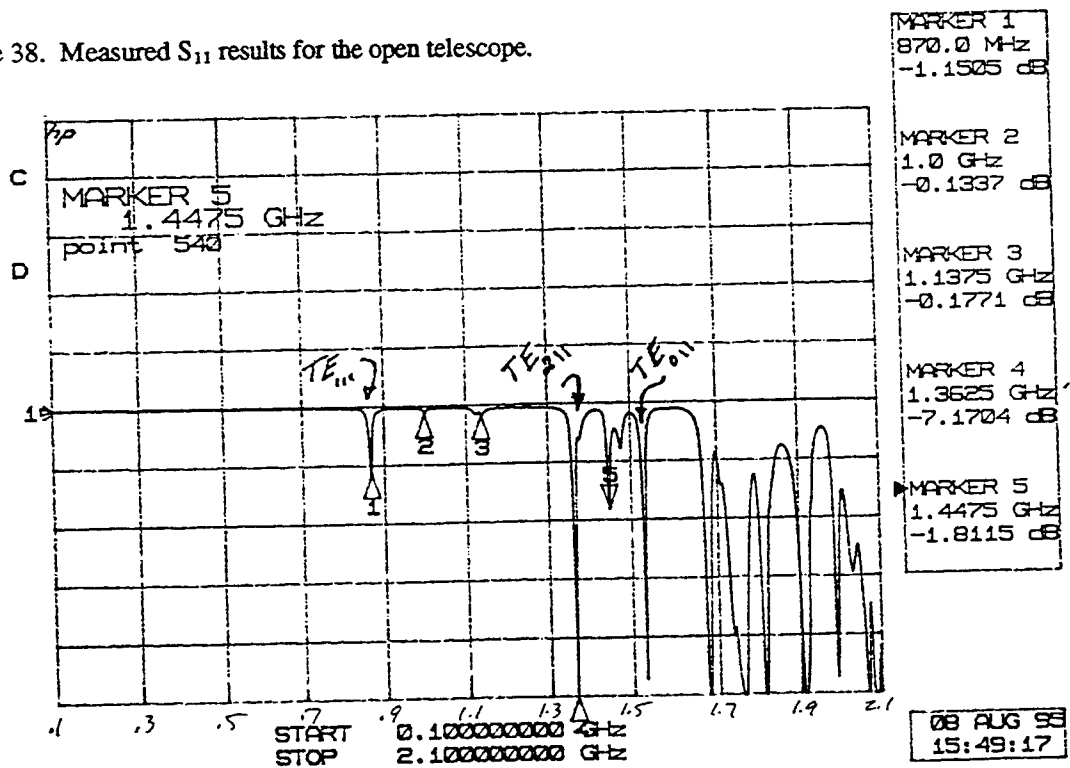


Figure 39. Measured S_{11} results for the covered telescope.

Further testing was done by Hugh Pohle over a frequency range of 1 GHz - 40 GHz. For these measurements, the telescope was excited by a transmitting antenna 40.0 inches in front of the aperture. A probe antenna was mounted at the opening to the eyepiece. The transmitting horn antenna was 3.02 inches in full length with aperture dimensions of 1.60 inches by 1.21 inches. The probe horn antenna was 0.88 inches in full length with aperture dimensions of 0.49 inches by 0.40 inches. Figure 40 shows one series of these measurements for 0.7 GHz - 18 GHz.

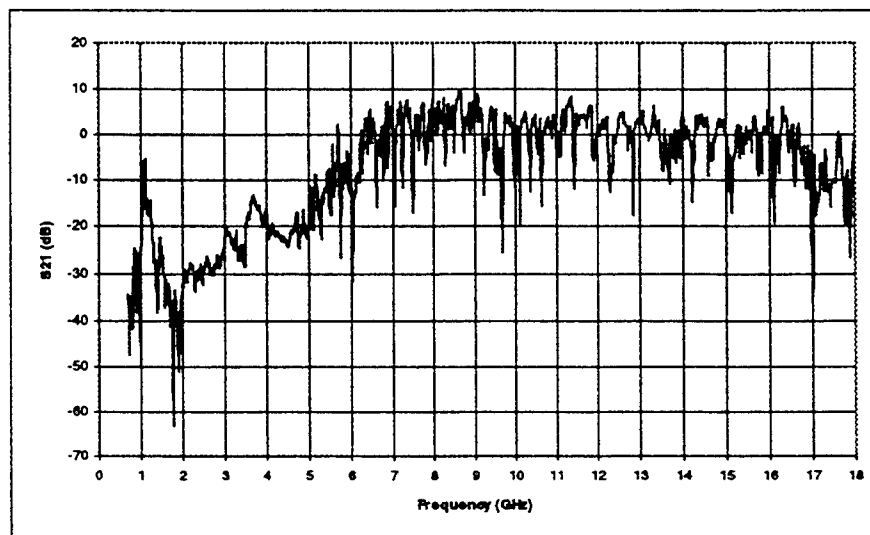


Figure 40. S_{21} measurements of the telescope.

CARLOS-3D Results vs. HFSS Results

Figures 41 -44 show the electric field distribution for the TE_{311} and TE_{211} modes for the open-ended cylindrical cavity at operating frequencies of 1.8 GHz and 1.35 GHz as given by CARLOS-3D. Comparing these figures to those of Figures 31 and 34, it is evident that the results obtained from CARLOS-3D corroborate the results obtained using HFSS. HFSS results have proven to be accurate as compared with the physical testing. This indicates that CARLOS-3D is functioning properly but does not provide the accuracy which is required for calculation of Q . However, implementing the accurate procedure outlined above for calculating the near fields will most likely provide the desired results.

Finally, the Q values for the two resonances at 1.8 GHz and 1.35 GHz are calculated by the HFSS post processor. These values are 55.5 at 1.8 GHz and 25.20 at 1.35 GHz. (Compare these values with Figures 14 which show Q on the order of hundreds and thousands.) It was necessary to use the post processor to calculate energy stored in the cavity along with power entering the cavity in order to calculate Q . The traditional method of calculating Q using the bandwidth of the resonances from the S_{11} data cannot be used because the resonances are not deep enough (see Figure 37).

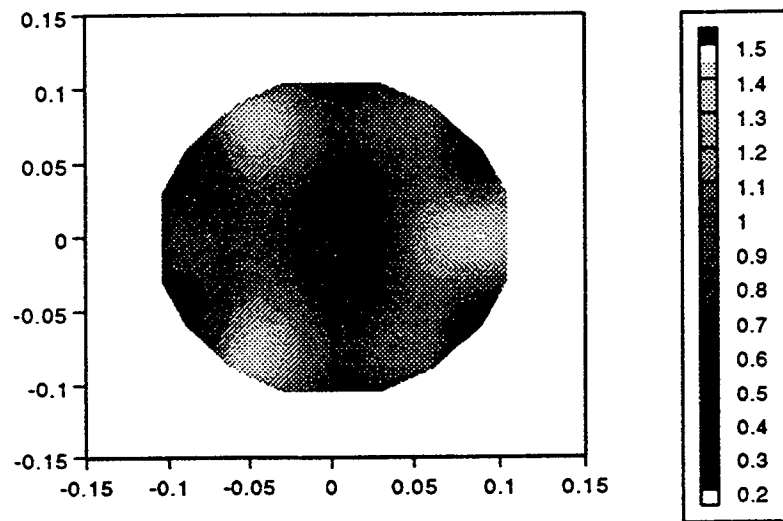


Figure 41. Magnitude of the Electric Field for the TE_{311} mode, $f=1.8$ GHz, $\phi=0^\circ$, $\theta=180^\circ$, $z=16.4$ cm

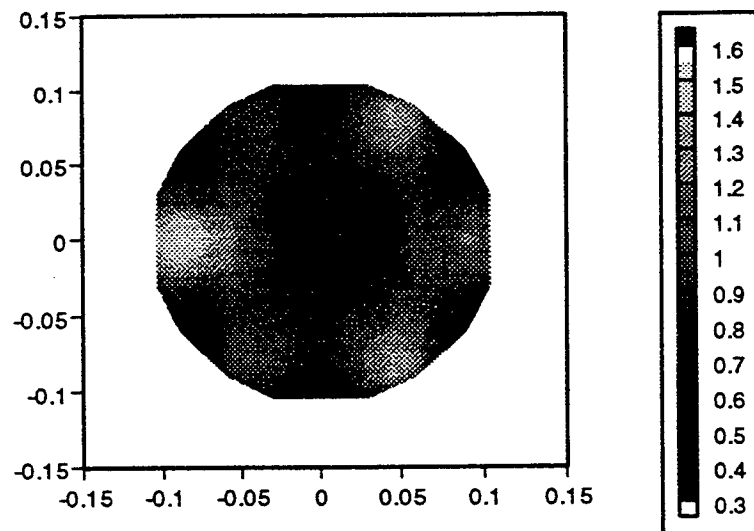


Figure 42. Magnitude of the Electric Field for the TE_{311} , $f=1.8$ GHz, $\phi=0^\circ$, $\theta=180^\circ$, $z=11.9$ cm

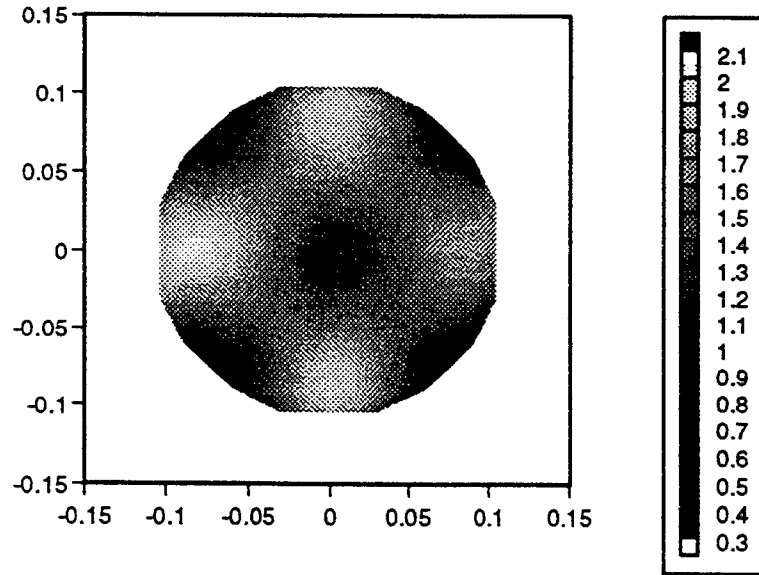


Figure 43. Magnitude of the Electric Field for the TE_{211} mode, $f=1.35$ GHz, $\phi=0^\circ$, $\theta=180^\circ$, $z=16.4$ cm

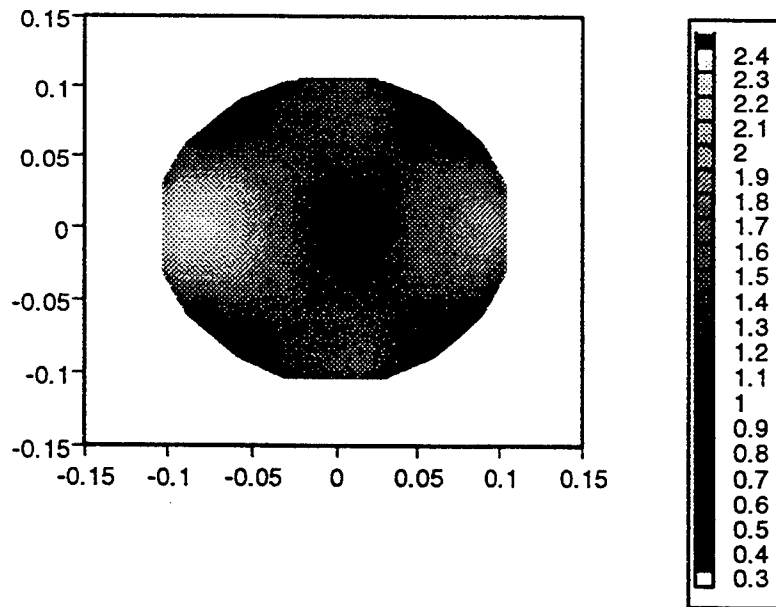


Figure 44. Magnitude of the Electric Field for the TE_{211} mode, $f=1.35$ GHz, $\phi=0^\circ$, $\theta=180^\circ$, $z=11.9$ cm

Conclusions

We implemented the method of iterative refinement in order to avoid inconsistencies in the system of linear equations, manifested in the Method of Moments, which translate to numerical instabilities. Iterative refinement is also implemented to restore full machine precision and to lower the condition number of the impedance matrix generated by MOM. We observed that for very dense meshes at lower frequencies, the condition number of the matrix was very large and in some cases identical to the machine floating point number. Under these conditions,

round off errors are highly plausible and hence the method of iterative refinement was implemented to rectify this problem. Having low condition number, however, does not automatically guarantee accuracy. We have observed that using the method of iterative refinement did not make noticeable improvements in the near field calculations. However, all the methods in this report were applied to an empty open-ended cylindrical cavity. Therefore, we may or may not see different responses to the method of iterative refinement related to changes in the internal geometry. Certainly the trend is for higher condition numbers and larger meshes. A full up model of this telescope may result in improvement of the solution using iterative refinement.

Higher order Gaussian Quadrature values are already implemented in the code for CARLOS-3D and the user has the opportunity to make the necessary changes to these numbers in order to improve the accuracy in calculating the electric field integral equation. Using improved Gaussian Quadrature did make substantial differences in the results. In many situations one does not have the luxury to invest large amount of time to search for suitable values. Therefore, an adaptive method of is necessary in order to avoid comprehensive analysis trying to find the appropriate order of the Gaussian Quadrature.

It has long been known that the currents at edges of a geometrical model behave differently than other parts of the model. Therefore, appropriate design parameters should be taken into account when modeling these edge currents. In the case of the open-ended cylindrical cavity, two edges (knife edge at the aperture along with a rectangular edge at the closed end) were modeled using non-uniform discretization to account for their behavior. Using non-uniform meshing schemes made a substantial improvement in the energy calculations. However, the same improvement was not noticed for the power calculation. The initial strategy was to implement more grid points in the aperture in order to improve the accuracy of the power calculation. However, a denser field mesh in the aperture did not improve the power calculations. This indicates that using the crude approximation for calculating the near fields severely affects the result of calculations and hence using a denser mesh in the aperture does not improve the results. Until the appropriate formulation for calculating the near fields, to include the variations of current over individual triangular cells and of distance to the field point is implemented in CARLOS-3D, adequate accuracy of near fields can not be obtained. CARLOS-3D calculations in this report were done for an empty open-ended cylindrical cavity. This was done solely to understand the effect of the formulations implemented in CARLOS-3D along with avoiding computationally intensive problem. Once the necessary modifications of CARLOS-3D have been made, the true model of the telescope will be implemented and analyzed.

CARLOS-3D and HFSS calculations in this report were done for an empty open-ended cylindrical cavity. Memory (RAM) and disk storage limitations were a reality beyond a $\lambda/9$ discretization at 2 GHz. Once the interior geometry of the telescope is implemented, it would be necessary to increase these computational capabilities in order to continue with these calculations. This approach would get us above the 4.5 GHz frequency range which is

just above the cutoff frequency for the inner cylinder of the Celestron-8 Telescope. This is a critical frequency for this system because below this frequency there is significant attenuation, drastically reducing the amount of energy reaching the detector plane.

The concept of determining energy at a particular point or area in a system by doing the inverse (antenna) problem is an attractive one. The vast amount of information contained in the HFSS solution can be used to determine optimum incidence angles to couple the energy into the system, field quantities anywhere in the system, system resonances, quality factor for the system, etc. HFSS was used in this study to demonstrate the concept; perhaps other types of codes such as TEMAC-3D (Temporal Electromagnetic Analysis Code, a finite difference time domain (FDTD) code), FERM (Finite Element Radiation Model, a MOM code) and CARLOS-3D could also be used to solve the antenna problem. All of these codes are available to DOD authorized personnel and are designed (with the exception of CARLOS-3D) to do the radiation problem. FERM and TEMAC-3D will be installed on the University of Denver computer to explore the radiation problem. CARLOS-3D would also be a good candidate to do the radiation problem because of its ability to handle dielectrics as well as conductors. However, modifications to CARLOS-3D would have to be made to do the radiation problem.

During discussions with Dick Holland (an independent contractor working on projects for the Satellite Assessment Center SatAC) at the final review of the project described in [11], it became apparent that detailed numerical analysis of the Celestron-8 Telescope should extend up to around 8 GHz. (Beyond this, statistical studies would be most appropriate.) Clearly, the model of the telescope would be prohibitively large for this high frequency. Another approach has to be considered - use of the SP2 at the Maui High Performance Computing Center (MHPCC). The SP2 is a multiprocessor computing network where parallelized software can take full advantage of the computational resources. The SatAC is currently working on producing a parallelized version of CARLOS-3D. Preliminary discussions with David Shnidman of MIT Lincoln Labs indicate that FERM can also be parallelized. This code, while it can't model dielectric regions, will do the radiation problem, something CARLOS-3D can not yet do. A full model of the Celestron-8 Telescope with the exception of the dielectric regions would still be very useful to study.

Any results from parallelized codes of CARLOS-3D and FERM run on the SP2 should be compared with results of the already validated codes run on a single high capacity computer. They should also be compared with results from an HFSS model. TEMAC-3D, which probably could be parallelized, is not projected to be done in the near future. It is understood from discussions with Sydney Blocher, PL/WSQ, that the process of parallelizing TEMAC-3D is more difficult due to the different types of boundary conditions. It would be useful to run this code for the radiation problem on a single processor machine to investigate its usefulness and accuracy when compared to the FEM of HFSS and MOM of FERM and CARLOS-3D.

Further progress at the University of Denver

An upgrade the University of Denver Engineering Department's IBM RS6000 PowerServer 570 with a 4.5 GByte hard disk and 384 MBytes of RAM (512 MBytes total) has been accomplished. The extra RAM was needed for the larger, more comprehensive models of the Celestron-8 Telescope. The extra drive was needed for data files, more software (see above) and for storage of temporary files during computer runs. We have successfully run the models telehf.pjt and telescop.pjt at 5 GHz and are currently doing frequency sweeps on both models in the range of 4.44 GHz to 5.5 GHz. Preliminary results from the telehf.pjt model indicate resonant frequencies consistent with a closed cylinder of the same size as the outside cylinder of the telescope. Run times are 6 hours per frequency for telehf.pjt and 13 hours per frequency for the telescop.pjt model.

We also plan on using the upgraded RS6000 to run CARLOS-3D, FERM and TEMAC-3D codes to model the Celestron-8 Telescope at approximately 5 GHz. CARLOS-3D, TEMAC-3D and HFSS will be used to look at the fields inside the cavity. HFSS, FERM and TEMAC-3D will be used to give information on the radiation problem; that is, radiation plots and directive gain at frequencies identified as "resonance frequencies" from the one port scattering parameter data. Familiarity with all of these codes will help determine which method (MOM, FEM or FDTD) and which pre- and post-processing features are best suited for these types of problems.

We have applied for and received 400 hours of computer time at the MHPCC. We will use this time to validate the parallelized software and to extend the frequency range of the Celestron-8 analysis to approximately 8 GHz.

Appendix A

It is known from experiment that the resonances of the cavities under consideration do not vary much from the closed empty cavity case. This knowledge is used to extend the traditional eigenvalue approach to a more general approach that is suitable for open-ended cavities. The method includes the following:

- 1) Solve the eigenvalue problem for Dirichlet and Neumann boundary conditions on the openings and Dirichlet boundary conditions elsewhere. This is the classic eigenvalue/eigenvector problem. From this step we can obtain estimates of eigenvalues and eigenvectors for the open-cylinder case. In a transmission line analogy, the Dirichlet conditions correspond to short-circuit loads, while the Neumann conditions correspond to open-circuits loads.
- 2) Use the eigenvectors obtained in step 1 as basis functions to generate an equivalent impedance matrix $Z_{int}(r, r')$ that fully describes the interior problem at any frequency.
- 3) Use GEMACS (General Electromagnetic Model for the Analysis of Complex Systems) to obtain an equivalent impedance matrix $Z_{out}(r, r')$ that fully describes the exterior problem at any frequency. We know that if the FD/FE (Finite Element) lattice extends to the far-field sphere, the impedance matrix is diagonal with terms equal to 376.7 ohms per square. The impedance matrix at the cavity opening(s) is not diagonal but dense, and the matrix terms vary extensively. However, by not having to work a FD/FE problem in the exterior, great computational savings can be made. GEMACS can be used to "bring the infinite boundary to the opening." This is a MOM/UTD (Unified Theory of Diffraction) hybrid problem (not an FD/FE problem. In the transmission line approach, this might be seen as referring a load impedance to another point along the transmission line via equations or smith chart. At the same time, GEMACS may bring a source or sources from their exterior locations to equivalent sources at the opening. The combination of impedance matrix and equivalent excitation completely describes the exterior physics to the interior problem.
- 4) Utilize Householder's method of modified matrices to combine the matrices together, enforcing E and H continuity at the open boundary. This is the technique used in GEMACS, by Harrington and Mautz, and by K. K. Mei. In the transmission line (TL) analogy it is like connecting two (or more) lines together at a single point. The act of "connection" ensures the transmission line voltage is the same on either side of the junction (analogous to E-field continuity) and the transmission line currents are equal and opposite on either side of the junction (analogous to H-field continuity).
- 5) Find the stationary points of the combined problem. At these points will be the complex resonant frequencies. From the complex resonant frequencies, obtain the resonances and Q's for each mode.

Depending on how rapidly the exterior equivalent impedance varies with frequency, this may require more than a matrix eigenvalue computation.

6) Utilize the eigenvectors of step 1 as basis functions for the coupling problem. Since these are not orthonormal for the open-cavity case, the Rayleigh-Ritz-Galerkin procedure will not yield a diagonal matrix. However, only the eigenvectors corresponding to eigenvalues near the operating frequency are required. Thus, the matrix problem to be solved for field coupling in the open-cavity case is quite small, perhaps no more than 20×20 .

REFERENCES

- [1] Wolfram, S., *Mathematica*, Addison-Wesley, 1991.
- [2] J.M. Putnam, L.N. Madgyesi-Mitschang and M.B. Gedera, CARLOS-3D™ Three Dimensional Method Of Moments Code, McDonnell Douglas Aerospace-East, December 1992.
- [3] General Electromagnetic Model for the Analysis of Complex Systems (GEMACS) - Version 5, Rome Air Development Center, AFSC, Griffiss Air Force Base, NY, December 1990.
- [4] R. R. DeLyser and P. Ensaf, "Analysis to Determine the Quality Factor of a Complex Cavity," Final Report for the Summer Faculty Research Program at the Phillips Laboratory, Sponsored by the Air Force Office of Scientific Research, September, 1994.
- [5] R. R. DeLyser, "An Analysis Approach to Determine Quality Factors of Large , Complex Cavities," Final Report for the Summer Faculty Research Program at the Phillips Laboratory, Sponsored by the Air Force Office of Scientific Research, September, 1993.
- [6] Balanis, C.A., *Antenna Theory*, John Wiley & Sons, New York, pp. 100-109, 1989.
- [7] Johnson J.H. Wang, *Generalized Moment Method In Electromagnetics*, John Wiley & Sons, New York, pp. 74-75, 1991.
- [8] Ishimaru, Akira, *Electromagnetic Wave Propagation, Radiation, and Scattering*, Prentice Hall, New Jersey, pp. 578-580, 1991.
- [9] Canning, Francis X. "Singular Value Decomposition of Integral Equations of EM and Applications to the Cavity Resonance Problem," *IEEE Trans, Antennas Propagat.*, vol. AP-37, no. 9, pp. 1156-1163, 1988.
- [10] William H. Press, Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling, *Numerical Recipes*, Cambridge University Press, New York, pp. 47-48, 1986
- [11] R. R. DeLyser and P. Ensaf, "Analysis to Determine the Quality Factor of a Celestron-8 Telescope," Final Report for the Summer Faculty Research Program at the Phillips Laboratory, Sponsored by the Air Force Office of Scientific Research, September, 1995.

Jean-Claude Diels report unavailable at time of publication.

A TOOL FOR THE FORMATION OF VARIABLE PARAMETER INVERSE
SYNTHETIC APERTURE RADAR
STACKED RANGE TIME INTENSITY, DOPPLER TIME INTENSITY,
AND WIDEBAND RANGE DOPPLER IMAGERY
OF ROTATING SPACE OBJECTS

James M. Henson
Assistant Professor
Department of Electrical Engineering

The University of Nevada
Reno, Nevada 89557

Final Report for:
Summer Research Extension Program
Phillips Laboratories

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Phillips Laboratories

December 1995

A TOOL FOR THE FORMATION OF VARIABLE PARAMETER INVERSE
SYNTHETIC APERTURE RADAR
STACKED RANGE TIME INTENSITY, DOPPLER TIME INTENSITY,
AND WIDEBAND RANGE DOPPLER IMAGERY
OF ROTATING SPACE OBJECTS

James M. Henson
Assistant Professor
Department of Electrical Engineering
University of Nevada

Abstract

The following paper describes the development and use of an Inverse Synthetic Aperture Radar (ISAR) signal and image simulation and processing software package (ISARTOOL) developed for Phillips Laboratories and funded through and Air Force Office of Scientific Research (AFOSR) Summer Research Program (SRP) Research Initiation Program (RIP) contract at the University of Nevada. It is the purpose of the program, ISARTOOL, to provide, using predicted RCS versus range data generated by the Lincoln Laboratories program ptdtable, techniques for generating variable parameter ISAR related imagery including Stacked Range Time Intensity, Stacked Doppler Time Intensity, and Wideband Range Doppler imagery. The program was developed under Builder Xcessory on a Silicon Graphics Indigo Workstation and exists in a fully interfaced point and click form. To date thirty (30) wideband data sets corresponding to three orientations of ten commercial satellites have been developed for use as program input data.

A TOOL FOR THE FORMATION OF VARIABLE PARAMETER INVERSE
SYNTHETIC APERTURE RADAR
STACKED RANGE TIME INTENSITY, DOPPLER TIME INTENSITY,
AND WIDEBAND RANGE DOPPLER IMAGERY
OF ROTATING SPACE OBJECTS

James M. Henson

1.0 Introduction

Much of the information in the present section can be found in [1] and [2]. It describes the types of raw and processed ISAR data and imagery available for subsequent processing and analysis. In the discussions that follow the terms wideband and narrowband refer to the bandwidth of the transmitted radar signal waveform and are somewhat relative. If a bandwidth of 1024 MHz is considered wideband, we might characterize a signal with one half or one fourth this bandwidth as narrowband. Clearly a single frequency is narrowband. We frequently address the problem in terms of range resolution and say that a bandwidth, B , which gives $\Delta R = \frac{C}{2B} = \frac{C\tau}{2}$ on the order of the dimension of the target is narrowband, while a B which gives ΔR much smaller than the object dimension is wideband. Obviously there are many intermediate bandwidths between 'narrow' and 'wide' band. Also recall that $h[n]$ and $H(f)$, the impulse and frequency response of an object are related through the Fourier Transform and that $h[n]$ represents a range (time) resolved signal.

For a single object model there are quite literally an infinite number of possible image sequences. Each is a function of orbital height and shape, elevation angle to target, object orientation and motion, and site parameters such as center frequency and bandwidth. Initially, we are creating 10 GHz (operating frequency), wideband (1024 MHz) HAYSTACK image sequences corresponding to four 360 degree great circle passes of the radar about each object at a fixed orientation. This wideband imagery / data will provide a portion of the input training data for future ASOI research and development. While the LL programs ptdtable, setupimagetbl, and imagetbl may be executed sequentially to generate an image sequence with range resolution $\Delta R = \frac{C}{2B} = \frac{C\tau}{2}$ and cross range resolution

$\Delta x = \lambda/2\theta$, it was desirable to develop algorithms and software capable of processing the raw data output (.wbtable) of the RCS prediction program, ptdtable. The reasons for this are several.

- 1) It is this data that must be processed to form wideband (narrowband) DTI and variable bandwidth RTI data and imagery.
- 2) Data corresponding to various range resolutions (bandwidths) can be generated far more efficiently and quickly from existing wideband .wbtable data than by repeatedly executing ptdtable for various resolution/bandwidth values.
- 3) Subsequent portions of LL code limits cross range image formation to angles of integration based on a 256 point DFT. It was necessary to extend the range of the DFT so that large angles of integration can be used for both linear and nonlinear focusing methods.

Having formed a wideband 360 degree .whtable data set for an object, the raw data is saved (by ptdtable) as magnitude/phase pairs in the following integer form.

$$m = 100 (A_{dB} + 200)$$

$$p = 100 (ArgA + 180)$$

Converting each data pair in the file according to $A = \text{Log}^{-1} \left[\frac{A_{dB}}{20} \right]$ where

$$A_{dB} = \frac{m}{100} - 200 \text{ and } ArgA = \frac{p}{100} - 180$$

We find the raw magnitude and phase values as

$$A = |x[j][k]|$$

$$ArgA = Arg(x[j][k])$$

where $x[j][k]$ is a complex valued two dimensional data set corresponding to the range versus aspect data.

The real and imaginary parts of $X[][]$ are found as

$\text{Re}\{x[][]\} = |X[][]| \cos(\text{Arg}X[][]) \text{ and } \text{Im}\{X[][]\} = |X[][]| \sin(\text{Arg}X[][])$. The data set shown in Figure 9 represents the raw ϕ versus f data after a DFT that converts the object frequency response data, $H(f)$, at each ϕ to the angle (ϕ) vs. range (R) data shown. The raw ϕ vs. f data can be obtained from $X[][]$ by applying a one dimensional inverse DFT along each row. If a DFT of the form

$$x[R_i][k] = \sum_{n=0}^{N-1} X[R_i][n] e^{-\frac{j2\pi n \left(k - \frac{N}{2}\right)}{N}}$$

is applied along each Column of the complex data set it is possible to form Wideband Range-Doppler Intensity and Phase imagery. Finally, we would note that the technique for manipulating the range resolution of .whtable data involves the coherent summing of complex data in the range dimension.

1.1 Range Doppler Data

As shown below, the raw data signal from a wideband radar is a continuous waveform due to a single target aspect. As the target moves with respect to the fixed radar position additional aspect sensitive signals, $V_i(t)$, are received. The data is processed to achieve resolution in the range dimension as shown in Figure 1.

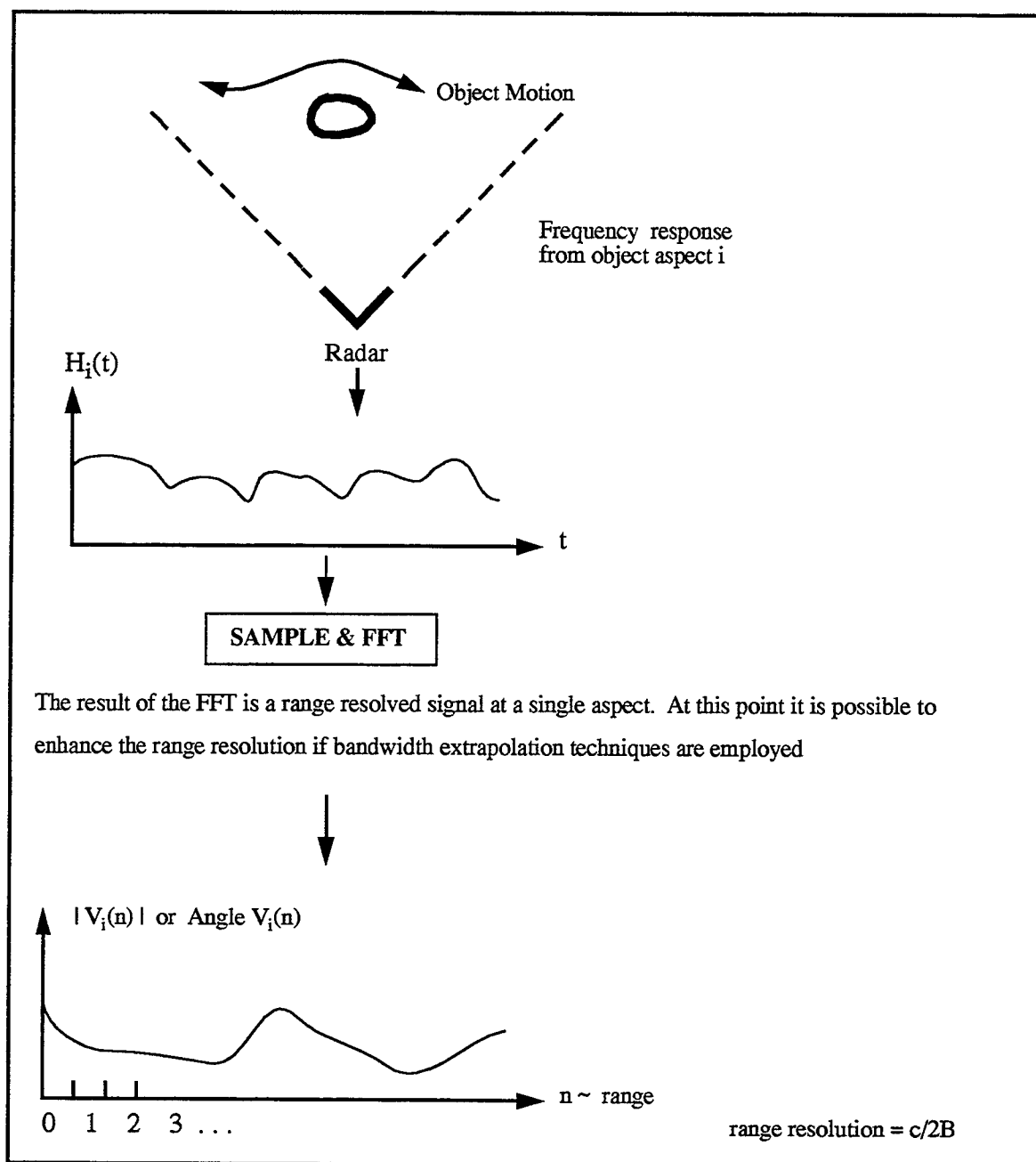


Figure 1

$V_i(n)$ may be thought of as the range 'signature' of the object when illuminated at aspect angle i .

This discrete signal can be displayed graphically as a function of n or as an image intensity strip. Discrete range resolved signals displayed in this fashion are called Range Time Intensity (RTI) strips.

As the object rotates some incremental amount, $\Delta\theta$, to aspect $i+1$, a second signal $V_{i+1}(n)$ may be generated. If these strips are stacked vertically, grey level intensity at a set of constant ranges may be observed in the vertical (aspect) dimension as shown schematically in Figure 2. Phase Time Intensity plots (strips) may also be created using phase rather than magnitude data.

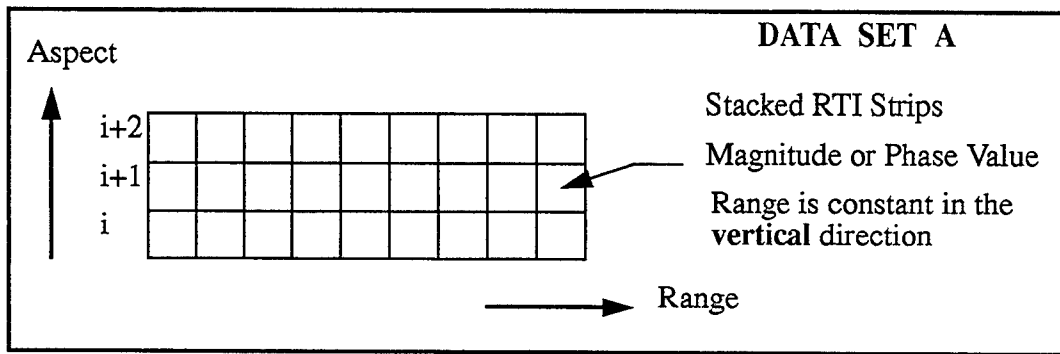


Figure 2

In this way a complex valued high resolution range signal is generated for each target aspect. The number of distinct angles available for further processing depends on the motion of the object as well as the true pulse width and PRF of the radar system used to collect the data.

Let us define this information, $V_i(n)$, $i = 1, 2, \dots, m$, as Data Set A. It is a complex set of values - one complex value for each range, R , and aspect angle, Θ

$$V_c(R, \Theta) = (a(R, \Theta) + jb(R, \Theta))$$

Now suppose that we repeat this data collection process using a Narrowband Pulse, that is, a pulse with

$$\Delta R = \frac{C}{2B} = \frac{C\tau}{2} \quad \text{on the order of the dimension of the object. Then at each aspect we will have only one return,}$$

$$V_i(t) = a \cos(\omega_c t + \Psi) \quad \text{where } \omega_c \text{ is the narrowband carrier frequency. The Transform of this signal is}$$

$$a\pi \left[e^{-j\Psi} \delta(\omega - \omega_c) + e^{j\Psi} \delta(\omega + \omega_c) \right]$$

If we 'stack' this data (intensity $a\pi$, or phase Ψ) in a **horizontal** fashion or plot it as a function of aspect angle, we

have the following data display possibilities:

- 1) intensity versus aspect strip images
- 2) intensity versus aspect plots (These are usually called RCS plots)
- 3) phase versus aspect strip images
- 4) phase versus aspect plots

Suppose that we now take the DFT of $i = 1, 2, \dots, N$ of these complex narrowband signal values, $V_i(\omega)$, where

$$N = \frac{\Theta_{Total}}{\Delta\Theta}$$

$$\Theta_{Total} \leq \sqrt{\frac{\lambda}{2R}}$$

$$\Delta\Theta \leq \frac{\lambda}{4R}$$

R = Radius of "Rotating" object

In this fashion one can create a series of Doppler intensity strips or plots as shown in **Figure 3**.

If these one dimensional doppler image strips (intensity or phase) are stacked vertically we form what are known as stacked Doppler Time Intensity (DTI) strips. Doppler Phase Strips can also be created if the phase data is used.

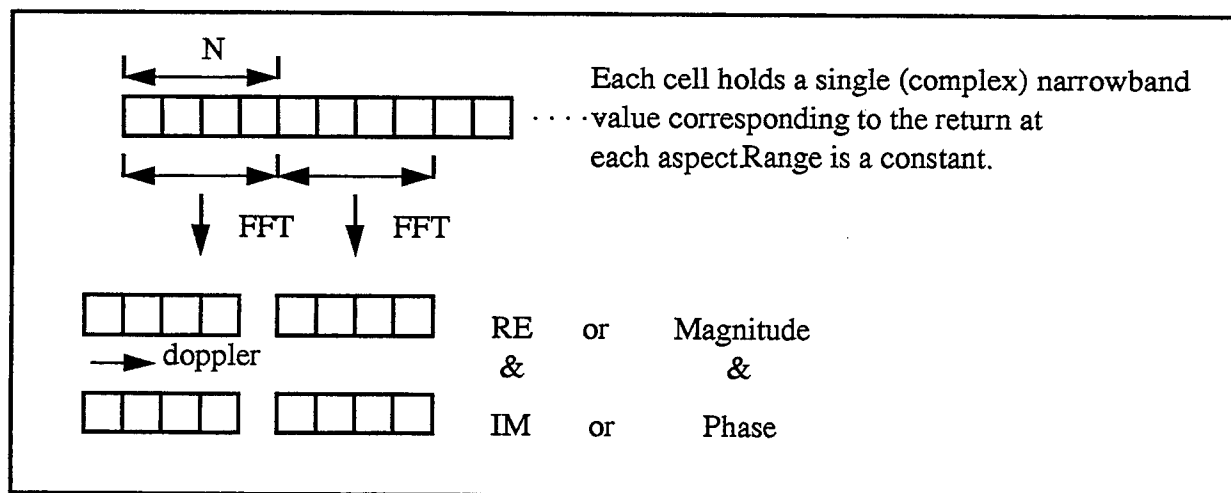


Figure 3

Each group of N values (in the horizontal direction) corresponds to a single cross range 'image' of the object. These images (intensity or phase) are one dimensional in the range dimension but resolved (linearly focused) in the Doppler or cross range direction.

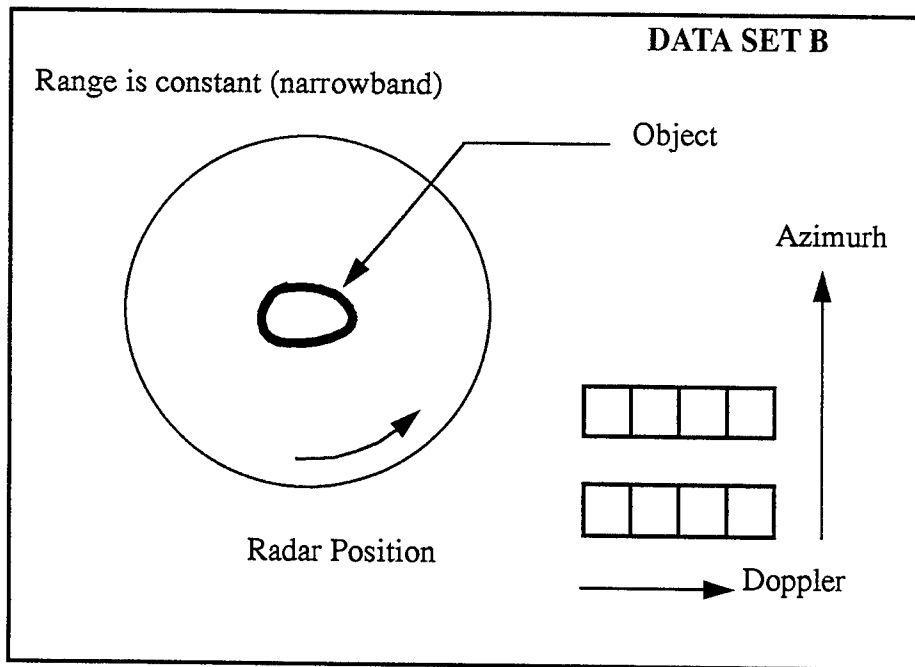


Figure 4

We define this information as Data Set B. It consists of a set of complex valued doppler (or cross range) arrays. Each one dimensional array corresponds to a particular object aspect. The data is depicted in **Figure 4**. Notice that the Narrowband DTI information of data set B can be created from data set A as shown in **Figure 5**.

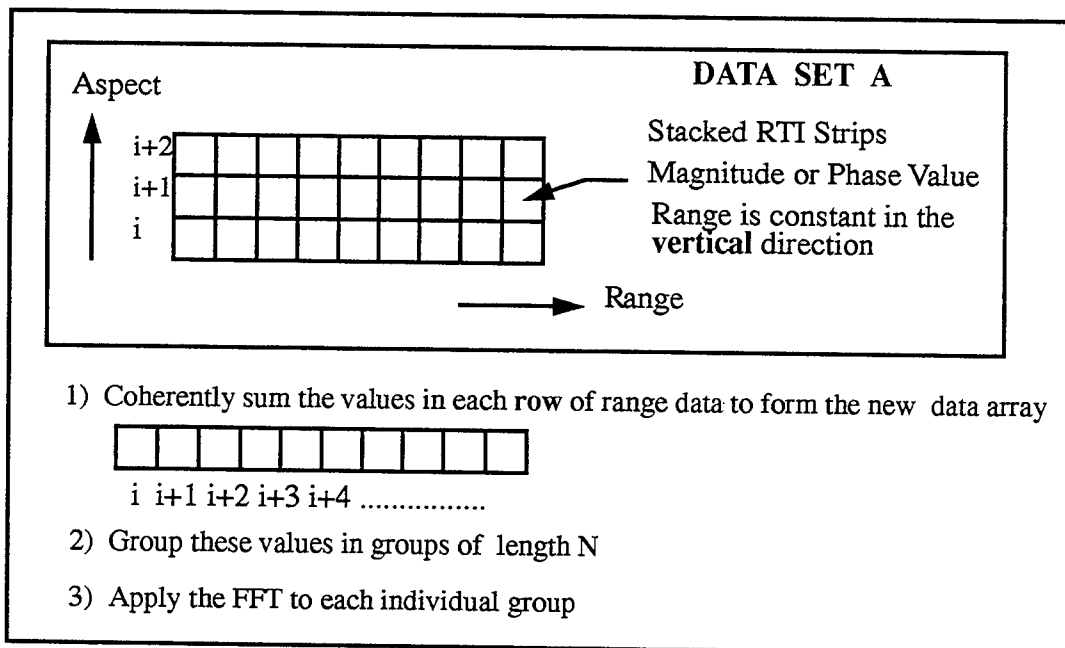


Figure 5

In other words, the narrowband information is recoverable from the wideband data. Actually, the narrowband and wideband range data represent the two range resolution extremes.

Finally, if the one dimensional FFT is applied to the columns of Data Set A (the complex wideband range resolved data) one can form true two dimensional range doppler (cross range) intensity and phase imagery. Note, however, that a single column of data may consist of range return data from any number of unique aspect increments, $i, i+1, \dots, i+M$, where the optimal increment is given theoretically as $\Delta\theta \leq \frac{\lambda}{4R}$. The optimal number of column elements used to form a cross range (doppler) image strip at a given range is $N = \frac{\theta_{Total}}{\Delta\theta}$, where $\theta_{Total} \leq \sqrt{\frac{\lambda}{2R}}$ is the angle of integration for a single image. Thus a total of $\frac{M+1}{N}$ images may be formed by this process. The basic processing is shown in Figure 6.

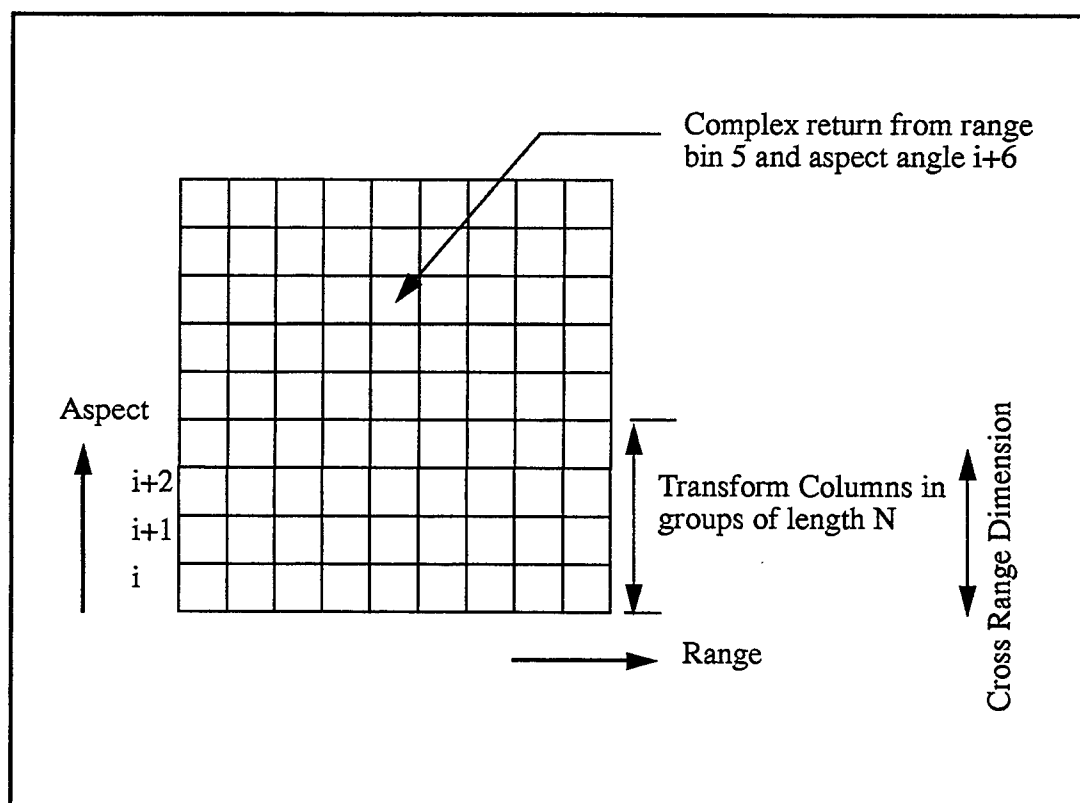


Figure 6

It should be noted that while the full wideband signal information content is available in the range Doppler Imagery, RTI and DTI strips and plots are expected to be of value in the satellite feature information extraction process.

2.0 ISARTOOL Description

It is the purpose of the ISARTOOL to provide the following image outputs utilizing as input Radar Cross Section (RCS) versus range data derived from the Lincoln Laboratories RCS prediction program ptdtable.

- 1) Stacked Range Time Intensity (RTI) Imagery
- 2) Stacked Doppler Time Intensity (DTI) Imagery
- 3) Linearly Focused Wideband RAnge Doppler (WBRD) Imagery
- 4) Wideband Doppler Time Intensity Imagery

In the discussions that follow reference is made to a number of user controlled variables in the software related to ISARTOOL as well as ptdtable. In Figure 7 the object on the turntable (x/y plane) is assumed to be stationary while the radar moves around the object along a circular arc in discrete angular increments of length $\Delta\theta = \frac{\lambda}{4R}$ radians or

$\Delta\theta = \frac{45\lambda}{\pi R}$ degrees. Lambda is the center frequency of the radar and R should be set to the known or estimated maximum radius of the object about its center of rotation (i.e. the z axis of the turntable). In actual processing terms the variable $\Delta\theta$ controls the processing interval or sampling rate.

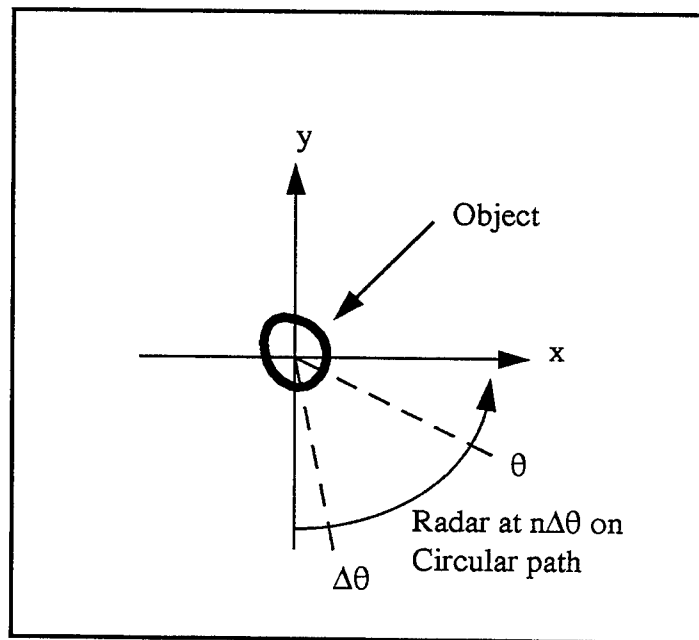


Figure 7

The value the user enters (in ptdtable) for 'number of samples' controls the maximum number of discrete increments of length $\Delta\theta$, over which the radar will be moved and available for processing in ISARTOOL. Note that moving the radar along a circular arc as shown in Figure 7 is entirely equivalent to holding the radar in a fixed position

and rotating the object about the z (turntable) axis..

If $\Delta\theta = 0.5$ degrees and $N=100$ samples, 50 range waveforms will be generated -- at viewing angles 0, 0.5, 1.0, 1.5, ..., 50 degrees. The variable θ shown in the figure is associated with the program `setupimagnetbl` where it is called the 'angle of integration'. NOTE that the **LL Users Guide to the RCS Toolbox (Version 3.0)** incorrectly defines the

angle of integration, θ , as $\Delta\theta$. The correct value (for optimal linear cross range focusing) is $\theta = \frac{180}{\pi} \sqrt{\frac{\lambda}{2R}}$

degrees. Together, θ and $\Delta\theta$ control the linearly focused cross range (or Doppler) resolution, $\Delta x \geq \sqrt{\frac{\lambda R}{2}}$. If the

angle of integration is taken such that $\theta < \sqrt{\frac{\lambda}{2R}}$, Δx increases correctly within the limits of the linear focusing

approximation. If $\theta > \sqrt{\frac{\lambda}{2R}}$, Δx theoretically decreases. However, the resulting imagery will be blurred (incorrectly focused) due to scattering centers 'walking' between resolution cells.

As shown in **Figure 8** the primary ISARTOOL window contains four buttons. Under the file button there are five options: Load wtable, Save Images, Save as TIFF, Wtable Information, and Exit.

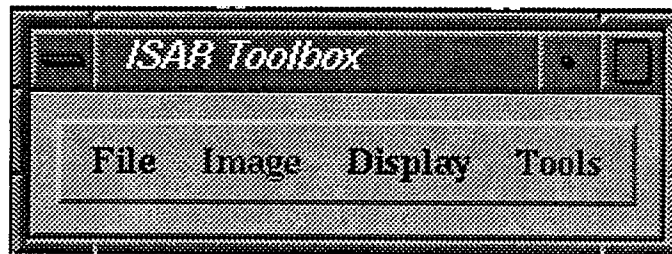


Figure 8

Figures 9 and 10 show the Load and Wtable Information windows. Note that data in the Wtable information window depends on the particular .wtable data set which has been loaded and cannot be modified in this window.

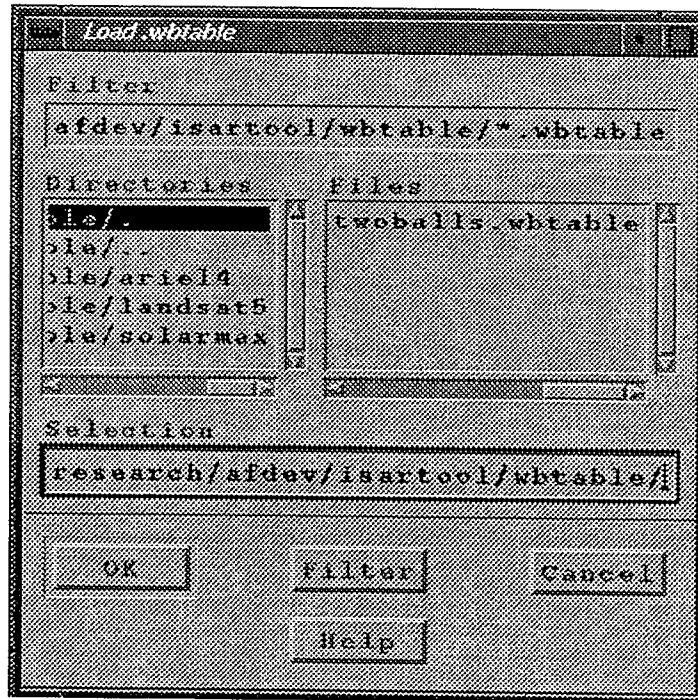


Figure 9

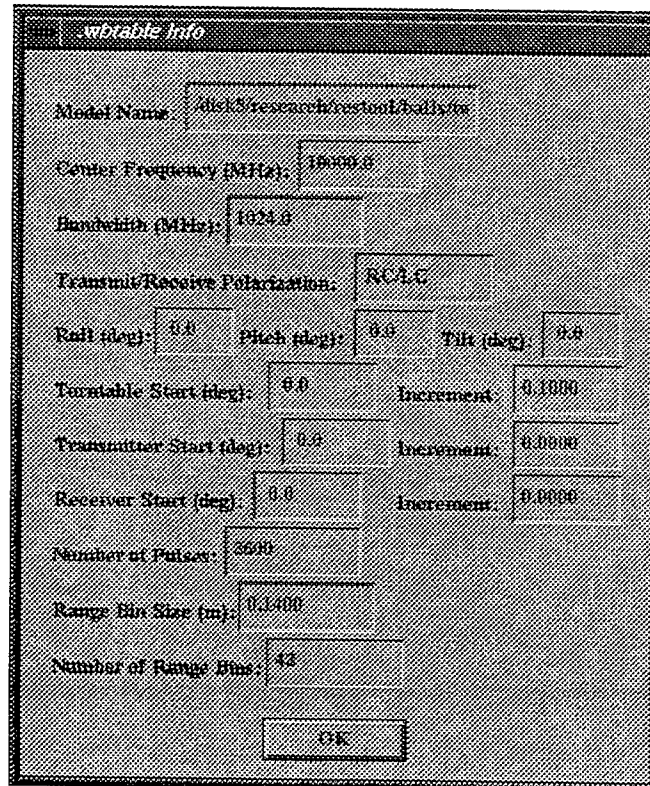
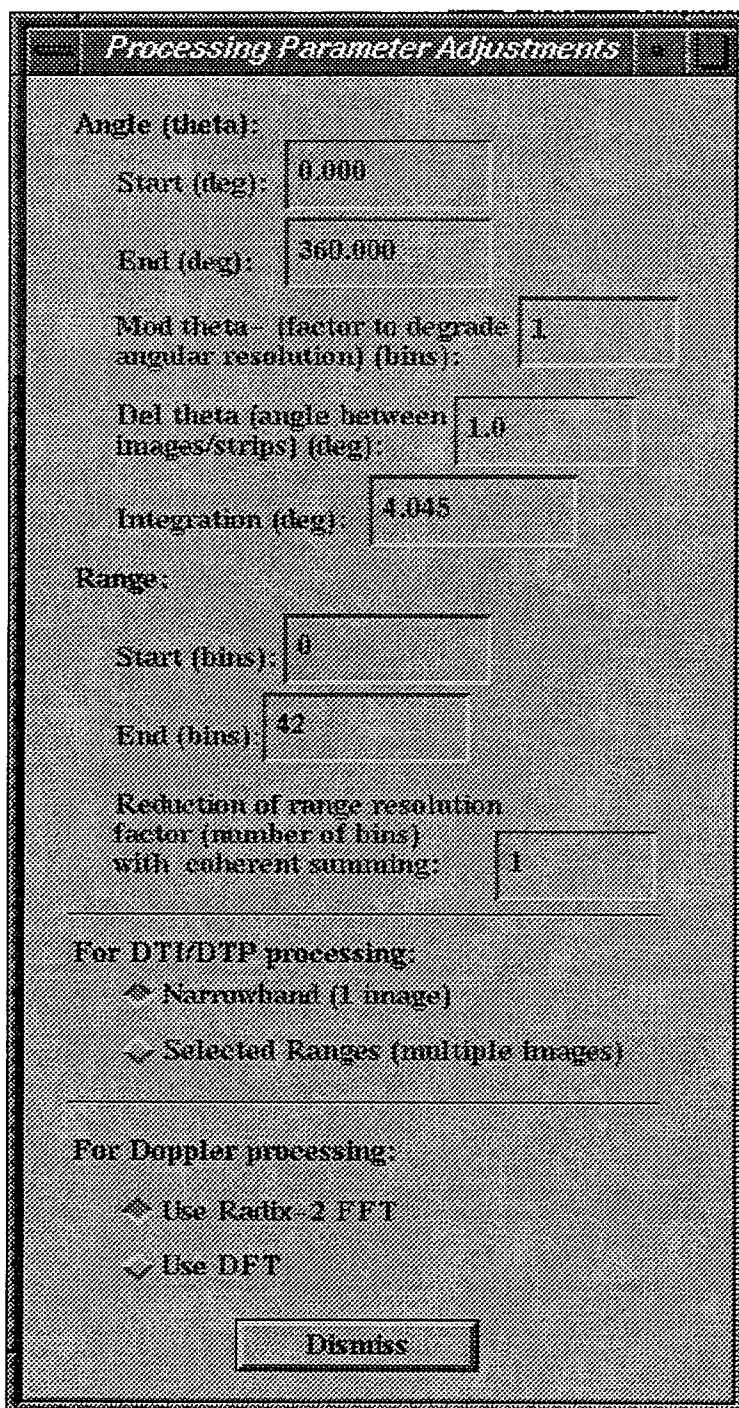


Figure 10

Under the Image button there are two pull down menu options: Image Parameter Adjustment and Create Images. The associated popup windows are shown in Figures 11 and 12.



The image shows a dialog box titled "Processing Parameter Adjustments". It contains several input fields and sections for parameter configuration.

Angle (theta):

- Start (deg): 0.000
- End (deg): 360.000
- Mod theta - (factor to degrade angular resolution) (bins): 1
- Del theta (angle between images/strips) (deg): 1.0
- Integration (deg): 4.045

Range:

- Start (bins): 0
- End (bins): 42
- Reduction of range resolution factor (number of bins) with coherent summing: 1

For DTI/DTP processing:

- ☒ Narrowband (1 image)
- ☒ Selected Ranges (multiple images)

For Doppler processing:

- ☒ Use Radix-2 FFT
- ☒ Use DFT

Dismiss

Figure 11

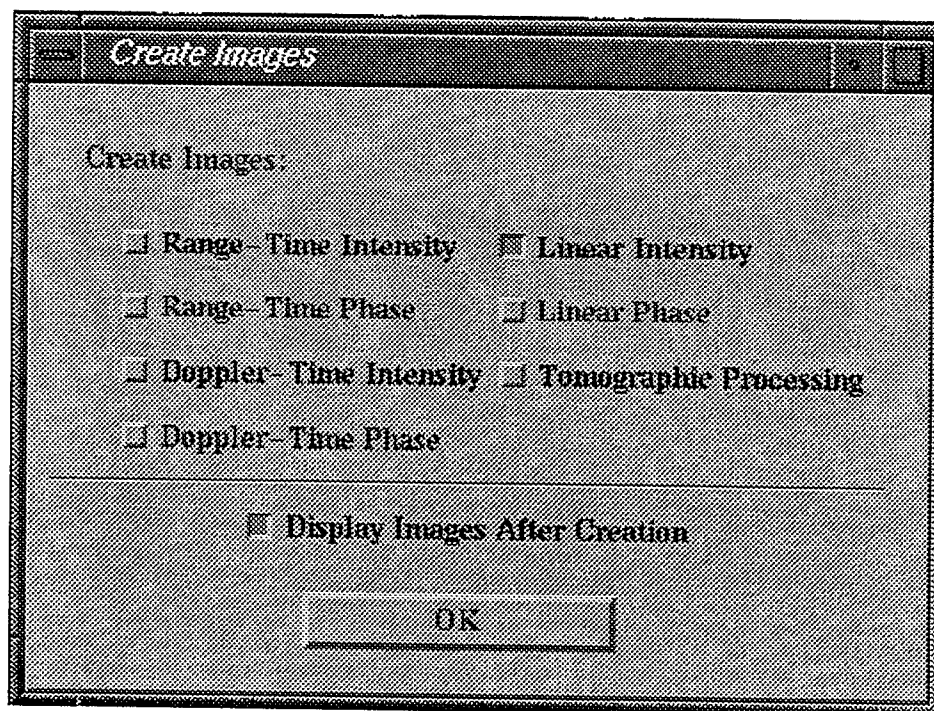


Figure 12

The Parameter Adjustment window contains a set of variables that can be used to control the formation of RTI, DTI, and WBRD imagery, while the create images window is used for determining which (set) of images the user wishes to create. The pull down menu under the Display button may be used to display any image that has been created as well as the Control Box shown in Figure 13.

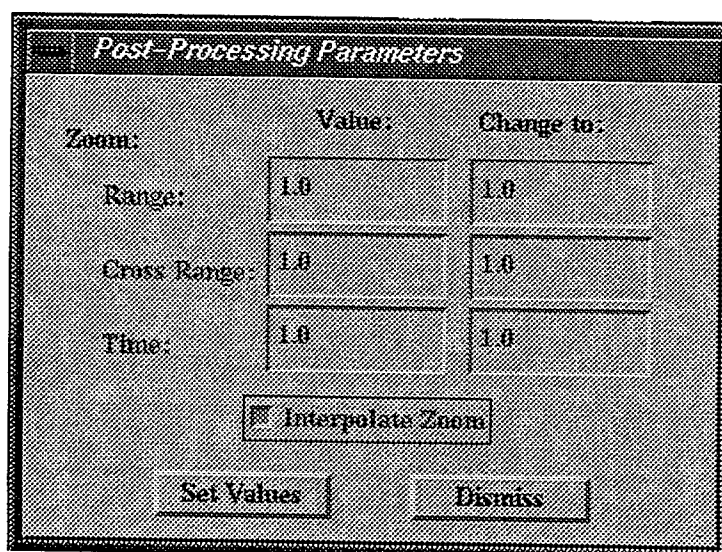


Figure 13

The Focused Image from RTI and DTI options under the Tools button are developmental and will not be discussed in the current report.

2.1 Range Time Intensity Image Formation

Stacked RTI imagery is generated directly from wtable intensity versus range data. Figure 14 shows the stacked wideband RTI image for a single sphere (point scattering center) rotating about a center not on the body.

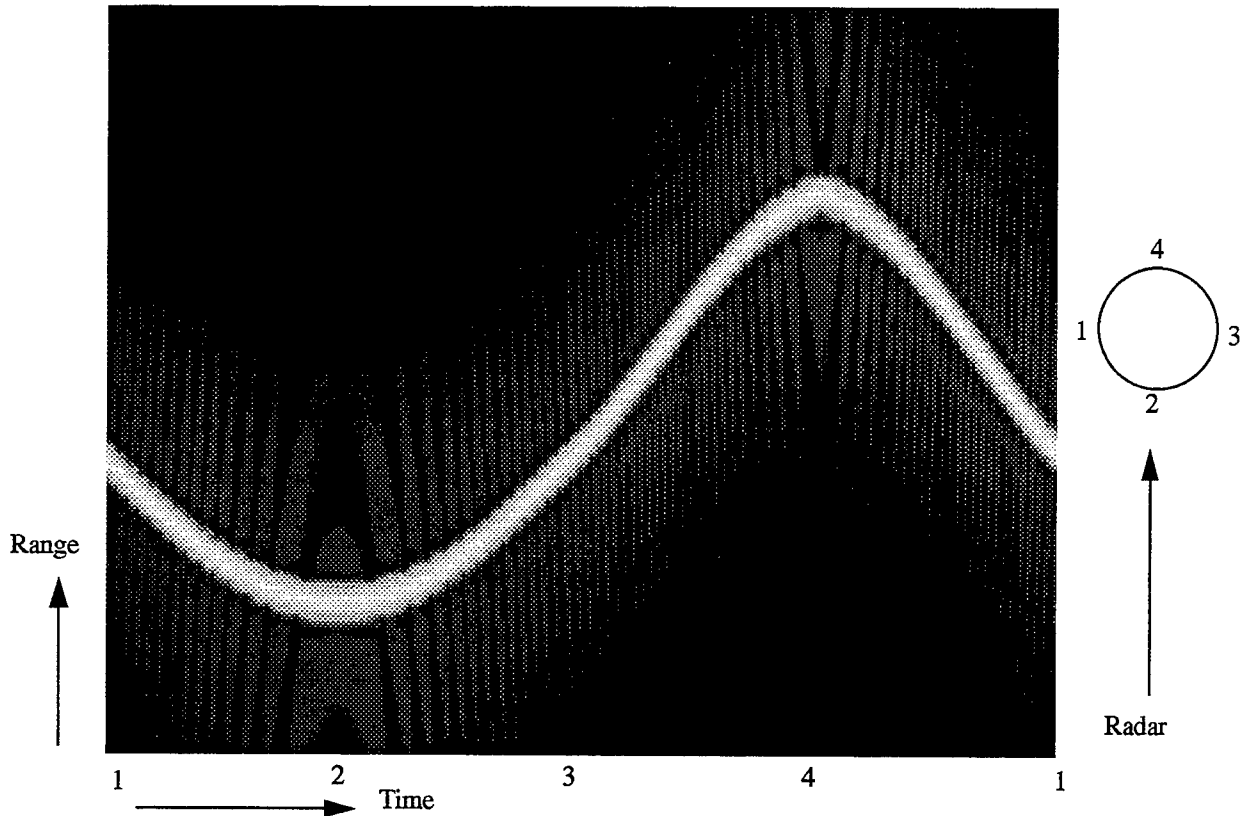


Figure 14

Parameters that may be adjusted to affect an RTI image (see Figure 4) include Start and Stop Angles, Del Theta, Range Start and Stop, and Range Resolution Reduction Factor. Del Theta, specified in real degrees, determines the angular increment between individual (vertical) image strips. If, for example, the RCS data was formed using a 0.1 degree turntable increment between pulses, setting Del Theta to 0.2 degrees will increase the angular increment to 0.2 degrees -- causing a relative reduction in the size of the RTI image in the time dimension. Similarly, decreasing the number of aspect angles (End Angle - Start Angle) will also result in a reduction of image extent in the time dimension. Decreasing the number of range bins (End Range - Start Range) will effect a corresponding reduction in image size in the vertical or range dimension. Finally, the Range Resolution Reduction Factor can be used to modify the original wtable range resolution. This is an integer value and determines the number of range bins that will be coherently summed to effect the new (decreased) resolution. Setting the reduction factor to two will cause a a reduc-

tion in the RTI image size in the range (vertical) dimension. Note that the variables Mod Theta and Integration Angle have no effect on stacked RTI imagery.

2.2 Stacked DTI Image Formation

Narrowband (single range bin) stacked Doppler Time Intensity data is formed via a coherent summation of all range bin data at each aspect angle followed by a series of N point DFTs in the (one dimensional) aspect dimension

where $N = \frac{\theta}{\Delta\theta}$, θ is the user specified Angle of Integration, and $\Delta\theta$ is the angular sampling increment. Figure

15 shows DTI image corresponding to the rotating sphere. Parameters that may be adjusted to generate a narrowband DTI image include Start and Stop Angle, Mod Theta, Del Theta, Integration Angle, and Range Start and Stop.

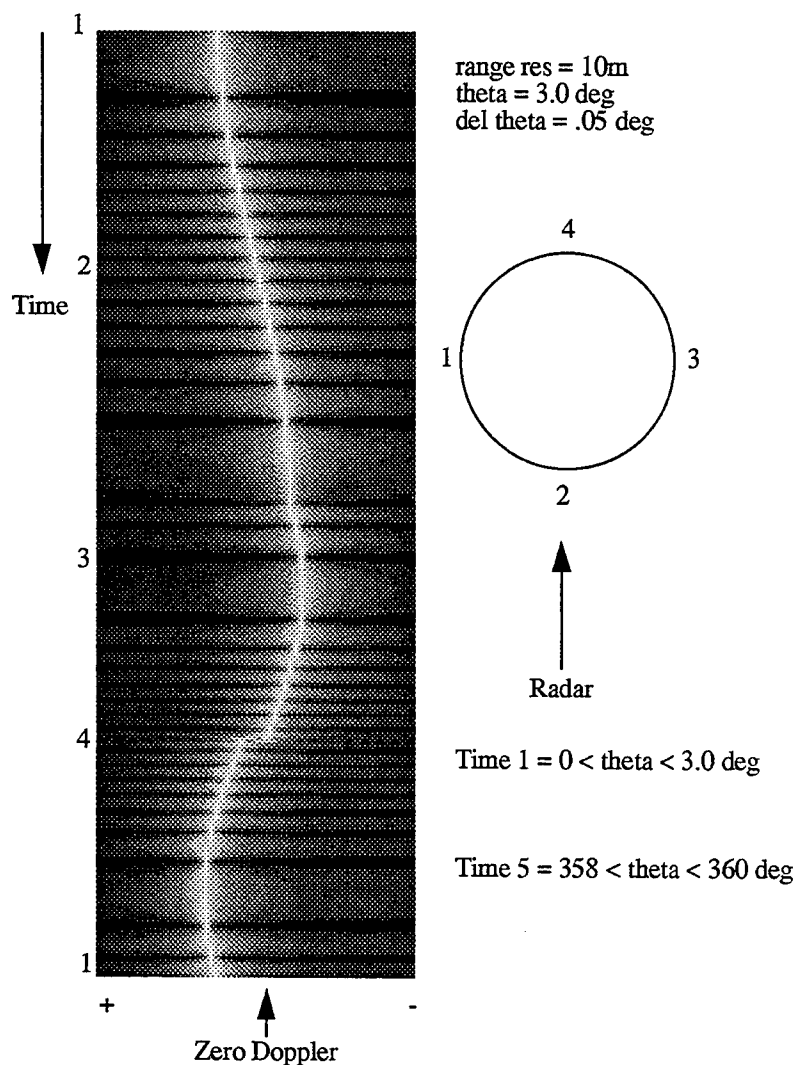


Figure 15

The values selected for Integration angle and Mod Theta, the angular increment between pulses, control the horizontal (cross range or Doppler) extent of the image. For example, for an Integration Angle of 10 degrees, and a Mod Theta of 1, the extent of the image in the horizontal dimension will depend on the angular increment (e.g. 0.1 degrees) found in the Wbtable Information Window. In this case the image extent will be $10/0.1 = 100$ pixels in the horizontal dimension. If Mod Theta is changed to 2, the Doppler extent becomes $10/2(0.1) = 50$ pixels. For a fixed Mod Theta increasing the Angle of Integration increases the Doppler extent of the DTI image. The variable Del Theta (in the case of DTI formation) controls the angle between individual (horizontal) image strips. If Del Theta and Integration Angle are set to the same value there will be no overlap of integration data. For example, if the integration angle is set to 10 degrees and Del Theta is set to 5 degrees, the first image strip represents a narrowband Doppler image over the aspect range 0 to 10 degrees. The second image strip represents a narrowband Doppler image over the aspect range 5 to 15 degrees, etc. The number of doppler strips control the vertical (time) extent of the DTI image. Thus for a fixed Start and Stop Angle, simultaneously increasing both the Integration and Del Theta will decrease the vertical extent of the DTI image.

If the Selected Ranges option under DTI/DTP Processing (Figure 4) is selected multiple DTI images can be formed. Each image in such a sequence represents a DTI image at a fixed range and fixed range resolution. The first image in the sequence corresponds to the DTI behavior of the closest range bin, while the final image corresponds to the DTI of the most distant range bin. Both the Range Resolution Reduction Factor and the Range Start and Stop parameters can be used in conjunction with this option. For example, if there are 40 range bins and the Resolution Reduction Factor is set to one, there will be 40 images in the sequence -- one for each range bin. If the Resolution Reduction Factor is increased to two, the number of range bins will be reduced to 20 and the sequence will consist of 20 DTI images. Modifying the values of Range Start and Stop can be used to limit the total range extent to any group of range bins of interest and will also decrease the number of DTI images in the sequence.

2.3 Linearly Focused Wideband Range Doppler Imagery

Wideband range doppler imagery, formed as described in Section 1 herein, can be generated (from wtable data sets) by ISARTOOL as a function of the following parameters: Start and Stop Angle, Mod Theta, Del Theta, Integration Angle, Range Start and Stop, and Range Resolution Reduction Factor. Figures 16 and 17 show WBRD imagery of the solarmax180 (180 is the orientation of the satellite on the turntable) with the following parameters: Integration Angle = 60 degrees, Del Theta = 60 degrees, Range resolution = 0.14 m, 72 range bins, Mod Theta =1 (with wtable angular sampling increment 0.05 degrees. Figure 16 represents the integration angle 60 to 119 degrees, while Figure 17 represents the integration angle 180 to 239 degrees. Figure 18 shows the result of a 360 degree integration.

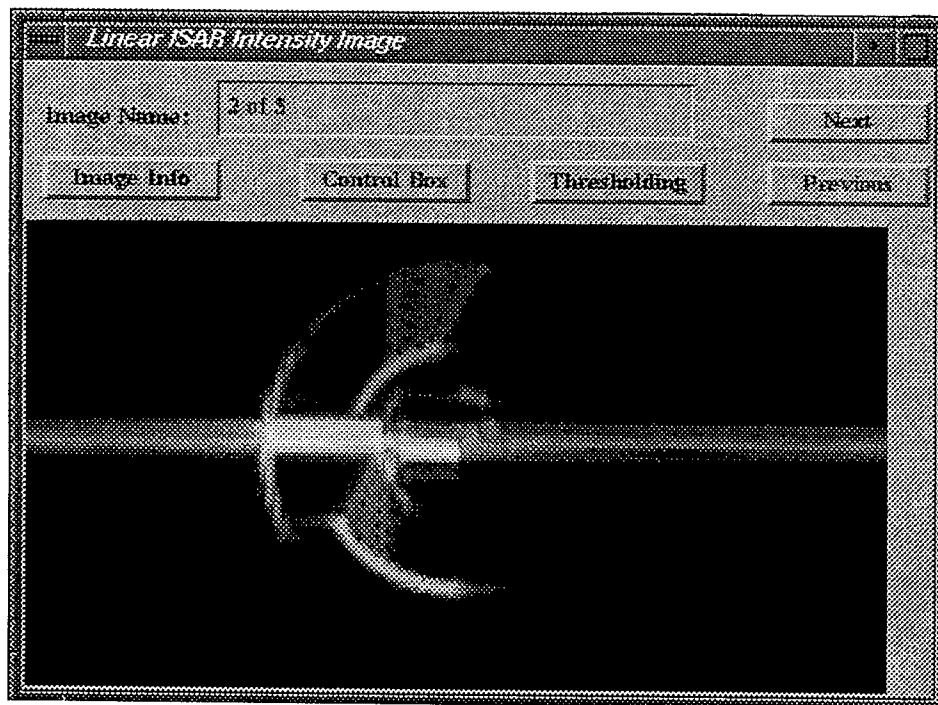


Figure 16

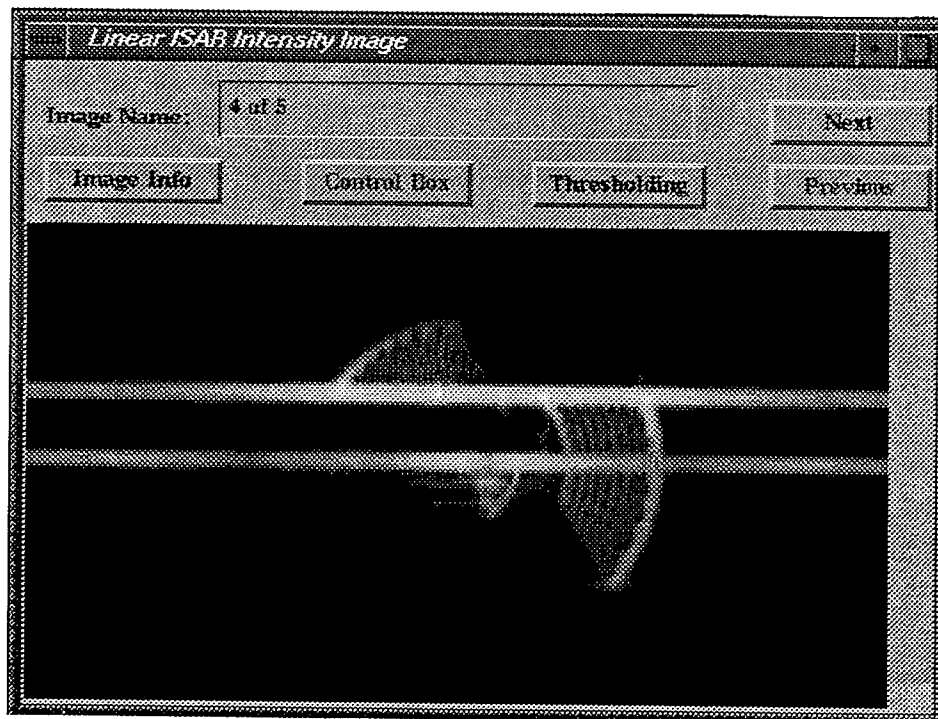


Figure 17

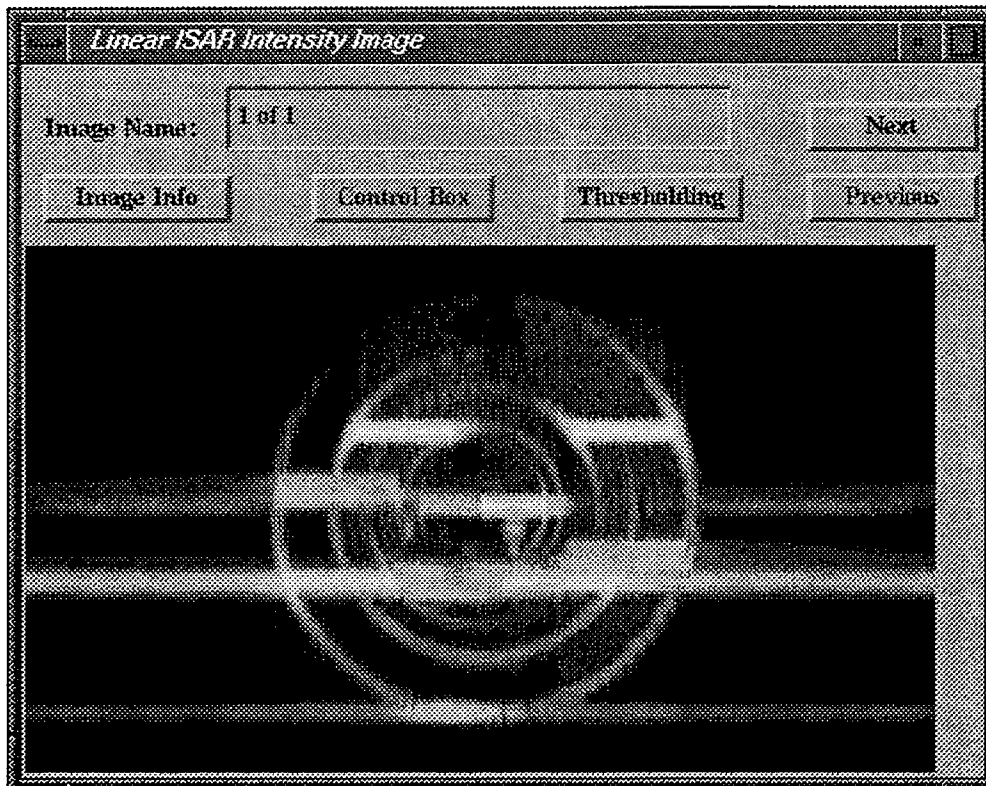


Figure 18

Note that resolution in the range (vertical) dimension is controlled by the signal bandwidth used to form the wtable data set as well as the range resolution reduction factor. Resolution in the cross range (horizontal) dimension is controlled by the integration angle, θ . That is, $\Delta x = \lambda/2\theta$. The number of images in a sequence depends on the Start and Stop Angle, Integration Angle, and Del Theta. A Start Angle of 0 degrees, a Stop Angle of 360 degrees, an Integration Angle of 36 degrees, and a Del Theta of 36 degrees will result in a sequence of 10 images. If Del Theta is decreased to 18 degrees (between images) 20 images will be generated at angles 0 to 36, 18 to 54, etc. As in DTI image formation Mod Theta controls the number of pulses (discrete aspect increments) within the angle of integration.

3.0 Model Construction and Examples

At this time ten models of commercial satellites have been constructed and include Ariel4, Ariel6, Explorer29, Landsat1, Landsat5, Nova4, Seasat, Solarmax, Tiros9, and TirosN. All models were constructed using the Lincoln Laboratories modeling tool ELIAS. RCS versus range data at three different orientations for each satellite was subsequently generated using the LL program ptdtable. Figure 19 shows the wire frame model of Solarmax which consists of 192 point scatterers (spheres), 17 boxes, 5 conics, and one rod. Figures 20 and 21 show a wideband (0.14m range resolution) stacked RTI image and a narrowband (one range bin) stacked DTI image of Solarmax180 respectively. The Angle of Integration used to form the DTI image was 3.126 degrees with angular resolution of 0.05 degrees.

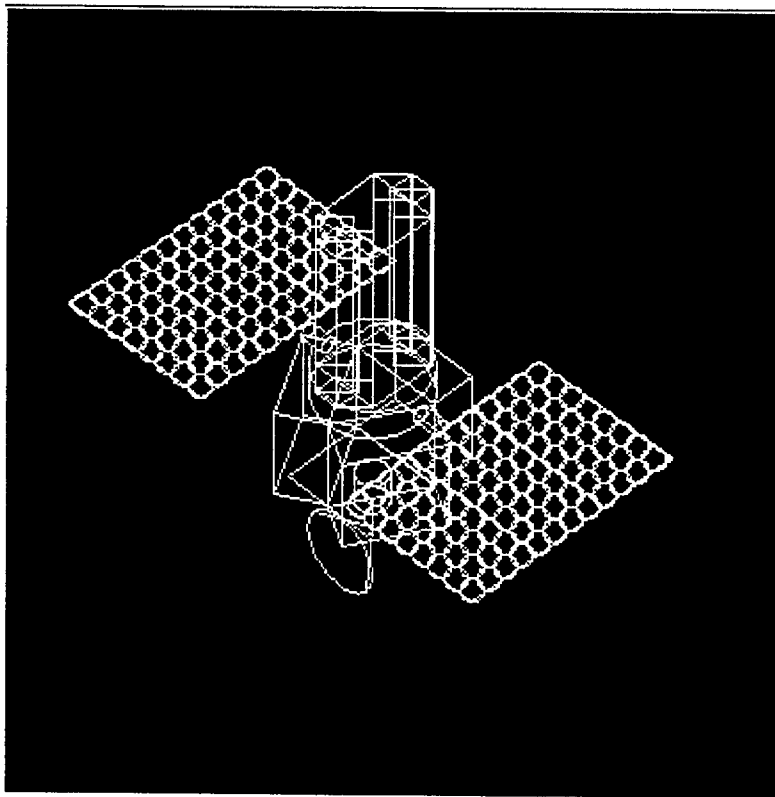


Figure 19

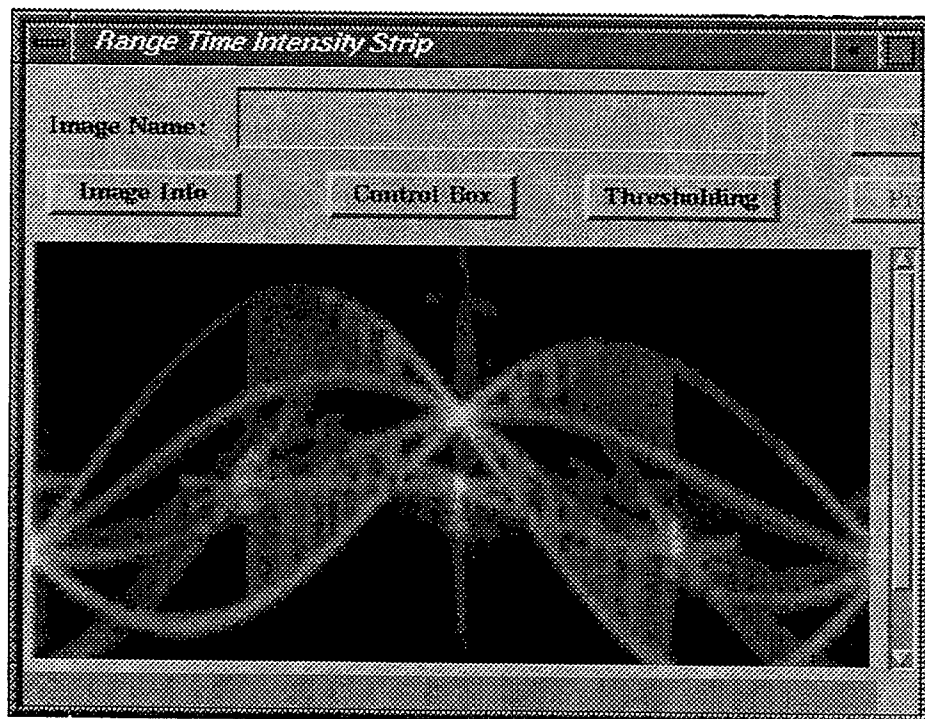


Figure 20



Figure 21

4.0 Summary

Given RCS versus range data sets as input ISARTOOL provides a compact tool for generating intensity and phase RTI, DTI, and variable bandwidth range Doppler image/data sets for rotating (orbiting) objects. All image results can be quickly and easily formed to simulate the performance of virtually any radar site and any imaging scenario. It is anticipated that ISARTOOL will prove to be invaluable for the development and analysis of automatic space object identification and assessment techniques. In addition to the effort described herein innovative signal processing techniques are also being developed to form two dimensional imagery from 1) narrowband (single range bin) DTI data sets and 2) variable bandwidth RTI data sets. Figures 22 and 23 show preliminary results of images from DTI only and RTI only data, respectively.

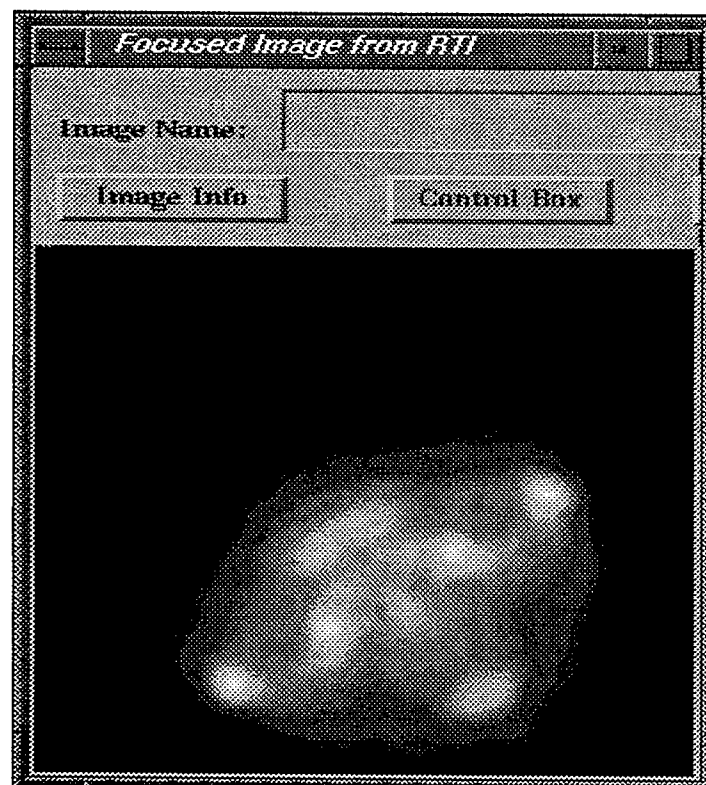
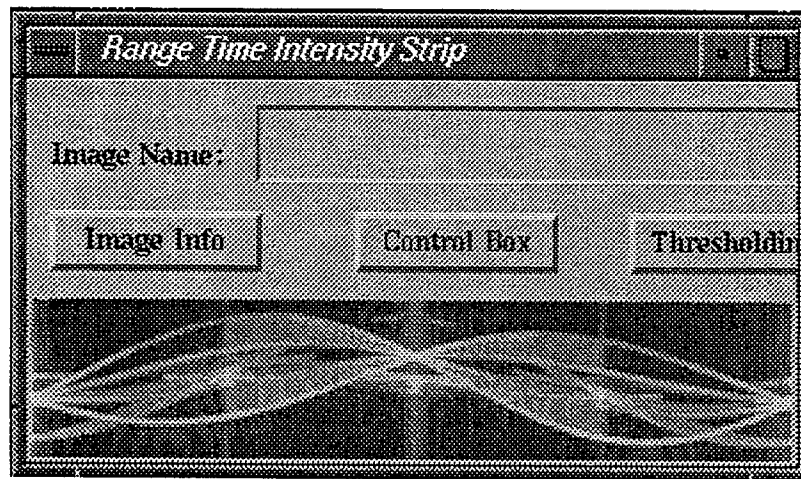


Figure 22

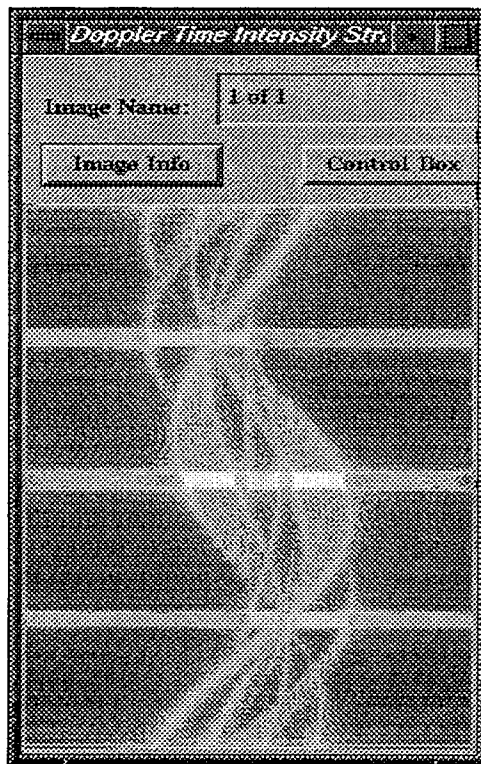


Figure 23

5.0 References

- [1] Ausherman, D.A., Kozma, A., Walker, J.L., Jones, H.M., and E.C. Poggio, Developments in Radar Imaging, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-20, No. 4, July 1984.
- [2] Brown, W.M., and R.J. Fredricks, Range-Doppler Imaging Motion Through Resolution Cells, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-5, No. 1, Jan., 1969.
- [3] Bramley, M., Burt, E., and K. Corbitt, User's Guide To RCS Toolbox, Version 3.0, MIT Lincoln Laboratories, Dec. 1992.
- [4] Carey, M.O., and C.M. DeRidder, Linear Periodogram Imaging, Project Report PSI-19, Lincoln Laboratories, April 1994.
- [5] Henson, J.M., An Introduction to Wideband Range and Doppler Cross Range Radar Imagery of Rotating Objects, Final Report, Air Force Office of Scientific Research, December 1993.
- [6] Mensa, D.L., High Resolution Radar Cross Section Imaging, Artech House, MA, 1991.
- [7] Mensa, D. and K. Vaccaro, Two-Dimensional RCS Image Focusing, *Proceedings of the 1987 Meeting of the Antenna Measurement Techniques Association*, Seattle.
- [8] Rihaczek, A.W., Principles of High Resolution Radar, Peninsula Publishing, Los Altos, CA, 1985.
- [9] Syracuse Research Corp., Coherent and Wideband Imaging Analysis Workbook, Prepared for USAF Air Training Command, 1979.
- [10] Walker, J.L., Range-Doppler Imaging of Rotating Objects, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-16, No. 1, Jan. 1980.

RADAR AMBIGUITY FUNCTIONALS

Gerald Kaiser
Professor
Department of Mathematical Sciences

University of Massachusetts at Lowell
1 University Avenue
Lowell, MA 01854

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research, Bolling Air Force Base, DC
and
University of Massachusetts at Lowell, Lowell, MA

December, 1995

RADAR AMBIGUITY FUNCTIONALS

Gerald Kaiser

Professor

Department of Mathematical Sciences

University of Massachusetts at Lowell

Lowell, MA 01854

email: kaiserg@woods.uml.edu

Abstract

Radar ambiguity functionals generalize the idea of ordinary ambiguity functions. The transmitting antenna, the target, and the receiving antenna are all allowed to execute independent and general (possibly nonlinear or accelerating) motions. The analysis takes place entirely in the four-dimensional space-time domain and is based on first principles starting with the emission, propagation, reflection, and reception of physical wavelets. The main limitation is that the antennas and the target are assumed to be point objects, and work is in progress to generalize the formalism to extended antennas and targets. A point antenna following an arbitrary trajectory $\alpha(t)$ in space defines an *emission operator* E_α , which transforms an arbitrary time signal $\psi(t)$ into a physical wavelet $(E_\alpha\psi)(\mathbf{x}, t)$. (That is, $E_\alpha\psi$ is a localized acoustic or electromagnetic wave, depending on the choice of the Green function in the definition of E_α .) This wavelet is reflected from a target following an arbitrary trajectory $\beta_0(t)$ in space, and the reflected wave is detected by the receiving antenna, which follows a third arbitrary trajectory $\gamma(t)$. To estimate the true target trajectory $\beta_0(t)$, the received signal $\chi(t)$ is matched with the computed signal $\psi_\beta(t)$ obtained by assuming a "trial" target trajectory $\beta(t)$. This matching of $\langle \psi_\beta, \chi \rangle$ defines the ambiguity functional $\tilde{\chi}[\beta]$, which depends on the trial trajectory $\beta(t)$ as a function. By maximizing the ambiguity functional (or minimizing the closely related *error functional*) an estimate is obtained for the true trajectory $\beta_0(t)$. When the receiving antenna is identical with the transmitting antenna and both are at rest (monostatic radar), and the target velocity is uniform in the line of sight, then the above ambiguity functional reduces to the usual wideband ambiguity functional, which is just the wavelet transform of the return with respect to the outgoing signal $\psi(t)$ as the basic wavelet. Upon the further assumption that ψ is narrowband this reduces to the narrowband ambiguity function, which is a windowed Fourier transform of the video signal of the return.

RADAR AMBIGUITY FUNCTIONALS

Gerald Kaiser

1. Introduction to Physical Wavelets

Physical wavelets were introduced in [1-5]. Their present generalization and application to radar was developed in [6-8]. They are "wavelets" in two distinct senses: In the old sense pioneered by Huygens, meaning localized acoustic or electromagnetic waves, and in the modern sense pioneered by Klauder, Morlet, Grossmann, Meyer, Mallat, Daubechies, and others [5,9], meaning functions that are all related by translation and scaling and that form a basis (or "frame") for a whole vector space of functions. In this section we give a brief introduction to such wavelets. For simplicity we specialize to *acoustic* wavelets, which are local scalar solutions of the wave equation in space-time. Electromagnetic wavelets are closely related but somewhat more complicated, since they take into account the vector nature of electric and magnetic fields; see [5], Chapter 9.

We use the following compact notation: The position vector in space \mathbf{R}^3 will be denoted by the boldface letter \mathbf{x} , and time will be denoted by τ . The point in four-dimensional space-time \mathbf{R}^4 with position vector \mathbf{x} and time coordinate τ will be denoted by the italic letter x , representing a "4-vector":

$$x = (\mathbf{x}, \tau).$$

The wave operator in space-time will be denoted by \square . It operates on functions $F(x) = F(\mathbf{x}, \tau)$ by

$$\square F(x) = \frac{1}{c^2} \frac{\partial^2}{\partial \tau^2} F(\mathbf{x}, \tau) - \nabla^2 F(\mathbf{x}, \tau),$$

where c is the constant propagation speed of the waves (sound or light) and ∇ is the gradient with respect to the space variables. A given scalar source function $J(\mathbf{x}, \tau)$ distributed in space-time generates an acoustic wave $F(\mathbf{x}, \tau)$ satisfying

$$\square F(x) = J(x).$$

The solution $F(x)$ is given by the four-dimensional convolution integral

$$F(x) = \int d^4 x' G(x - x') J(x'), \quad (1)$$

where G is the *retarded Green function*

$$G(x) = G(\mathbf{x}, \tau) = \frac{\delta(\tau - |\mathbf{x}|/c)}{4\pi|\mathbf{x}|}. \quad (2)$$

(For electromagnetic waves, G would be a dyadic Green function associated with Maxwell's equations.) Suppose now that we have a small transmitter, one that can be regarded as a point source, moving along an arbitrary trajectory in space given by $\mathbf{x} = \mathbf{r}(t)$. While moving along this trajectory, the transmitter is emitting a (real) signal $\psi(t)$. The trajectory of the transmitter in *space-time* \mathbf{R}^4 can be parameterized as a 4-vector function $\alpha : \mathbf{R} \rightarrow \mathbf{R}^4$ given by

$$\alpha(t) = (\mathbf{r}(t), t).$$

Since the source function is now concentrated along $\alpha(t)$, the resulting wave (1) reduces to

$$\begin{aligned} F(x) &= \int_{-\infty}^{\infty} dt G(x - \alpha(t)) \psi(t) \\ &= \int_{-\infty}^{\infty} dt G(\mathbf{x} - \mathbf{r}(t), \tau - t) \psi(t). \end{aligned} \quad (3)$$

Given the trajectory $\alpha(t)$, let us define the *emission operator* E_α as the operator transforming the time signal $\psi(t)$ into the space-time wave $F(x)$ given by (3):

$$(E_\alpha \psi)(x) = \int_{-\infty}^{\infty} dt G(x - \alpha(t)) \psi(t). \quad (4)$$

When the trajectory $\alpha(t)$ is *linear*, i.e., when the transmitter moves at a constant velocity through space, then $E_\alpha \psi$ turns out to be a translated and scaled version of $\psi(t)$ [6]. The simplest case occurs when the transmitter is at rest at the origin, so that $\mathbf{r}(t) = \mathbf{0}$. Then

$$(E_\alpha \psi)(\mathbf{x}, \tau) = \frac{\psi(\tau - \rho/c)}{4\pi\rho}, \quad \text{where } \rho = |\mathbf{x}|. \quad (5)$$

Since the right-hand side depends only on ρ and τ , wavelets emitted by stationary sources are spherical and can be easily visualized. One such example (for a particular choice of ψ associated with analyticity) is shown in Figure 1.

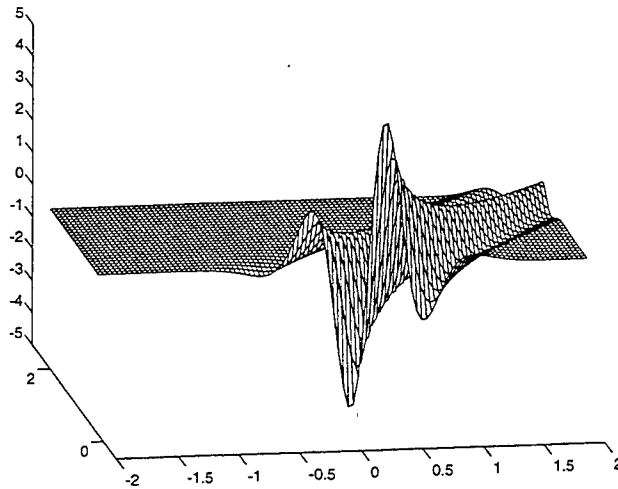


Figure 1. Example of an acoustic wavelet with stationary source as in Equation (5), plotted as a function of time τ (left to right) and radius ρ . This is the emitted part of the wavelet in Figure 11.6 of [5], given by Equation (11.41) with $\alpha = 10$.

Equation (5) states the intuitively obvious fact that the wave received at x at time τ is a delayed (by ρ/c) and attenuated (by $1/4\pi\rho$) version of the transmitted signal. When the transmitter is moving at a constant nonzero velocity, the resulting wave furthermore exhibits a Doppler effect, which amounts to a *scaling* along the direction of motion. Since translations (like delays) and scalings form the “raw material” for the standard wavelets [5,9], we call $E_\alpha\psi$ a *generalized acoustic wavelet*. It is “generalized” because for nonlinear $\alpha(t)$, $E_\alpha\psi$ can no longer be described simply as a translated and scaled version of ψ but is still given by the integral (4).

The process of reception works in the opposite direction: A receiver moving along a trajectory $\alpha(t)$ converts an acoustic wave $F(x)$ into a function of time (the received signal) by “measuring” the wave along its trajectory. This can be formulated as a *reception operator* R_α defined by

$$(R_\alpha F)(t) = g_\alpha F(\alpha(t)), \quad (6)$$

where g_α is a gain factor associated with the receiver. R_α transforms functions of four variables to functions of one variable. This is only the simplest model for reception, but it will do for the present purpose.

2. Applications to Radar and Sonar

Suppose a transmitter is moving along a trajectory $\alpha(t)$ as above, emitting a time signal $\psi(t)$. The resulting wave $E_\alpha\psi$ is then *reflected* from a point “target” moving along a second trajectory $\beta(t)$, and the reflected wave is finally *received* by a receiver moving along a third trajectory $\gamma(t)$, giving a time signal $\chi(t)$. This situation is depicted schematically in Figure 2.

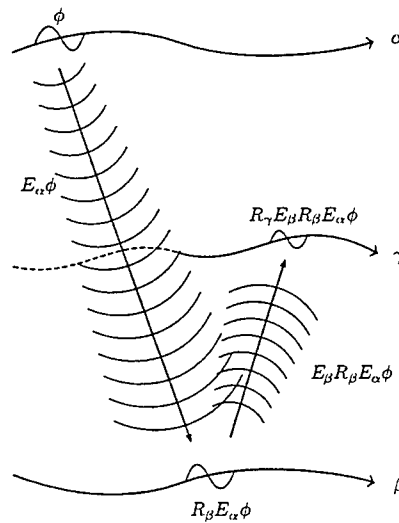


Figure 2. The signal ϕ is emitted from the trajectory α , reflected from β , and received at γ .

General Radar/Sonar problem: Given the trajectories $\alpha(t)$ and $\gamma(t)$ of the transmitter and the receiver, as well as the transmitted and received signals $\psi(t)$ and $\chi(t)$, estimate the trajectory $\beta(t)$ of the target.

Before attempting to solve this problem we must provide a model for the reflection of waves from a point target moving along the trajectory $\beta(t)$. Given an arbitrary wave $F(x)$ in space-time, let us model the reflected wave by assuming that $F(x)$ is first received by the target, then immediately re-emitted. That is, the target momentarily acts as a receiver, then as a transmitter. (In electromagnetics, this corresponds to formulating scattering by observing that the incident wave generates a current on the target, which then radiates the scattered wave.) According to (4) and (6), the reflected wave is then

$$F_{\text{refl}}(x) = (E_{\beta} R_{\beta} F)(x) = g_{\beta} \int_{-\infty}^{\infty} dt G(x - \beta(t)) F(\beta(t)). \quad (7)$$

Note that reflection is represented by the operator $E_{\beta} R_{\beta}$ converting a function of four variables (the incident wave) to another function of four variables (the reflected wave). In the present context, the "gain factor" g_{β} is interpreted as a *reflection coefficient*. With emission, reception, and reflection modeled by (4), (6), and (7), we are ready to tackle the radar/sonar problem. If we knew the target trajectory β (which we do not), then we could compute the received signal as

$$\psi_{\beta} = R_{\gamma} E_{\beta} R_{\beta} E_{\alpha} \psi, \quad (8)$$

where the dependence on the unknown trajectory β of the target has been made explicit while the dependence on the known trajectories α and γ of the transmitter and receiver is left implicit. Inserting the definitions of the emission and reception operators gives the detailed formula

$$\psi_{\beta}(t) = g_{\gamma} g_{\beta} \iint dt' dt'' G(\gamma(t) - \beta(t')) G(\beta(t') - \alpha(t'')) \psi(t''). \quad (9)$$

Our strategy for solving the radar/sonar problem is to compute ψ_{β} for a *trial trajectory* β , then *match* the result with the actual received signal $\chi(t)$. "Matching" here means simply taking the *inner product*, i.e., integrating over time:

$$\langle \psi_{\beta}, \chi \rangle = \int_{-\infty}^{\infty} dt \psi_{\beta}(t) \chi(t). \quad (10)$$

Equation (9) shows that the right-hand side of (10) depends on the entire *history* of the target motion.* Hence $\langle \psi_{\beta}, \chi \rangle$ depends on β as a *functional* rather than an ordinary function. We call it the *ambiguity functional* of β and denote it by $\tilde{\chi}[\beta]$:

$$\tilde{\chi}[\beta] = \langle \psi_{\beta}, \chi \rangle. \quad (11)$$

* To be precise, it depends on those values $\beta(t')$ for which (A) there are times t and t'' such that $\chi(t) \neq 0$ and $\psi(t'') \neq 0$, and (B) $\alpha(t'')$, $\beta(t')$ and $\gamma(t)$ are *causally ordered*, so that an acoustic or electromagnetic wave can travel from $\alpha(t'')$ to $\beta(t')$ to $\gamma(t)$ at the speed c . This causality is built into the Green functions in (9).

The Schwarz inequality states that

$$|\langle \psi_\beta, \chi \rangle| \leq \|\psi_\beta\| \|\chi\|, \quad (12)$$

where $\|\psi_\beta\|$ and $\|\chi\|$ are the norms of ψ_β and χ , i.e., their squares are the energies:

$$\|\psi_\beta\|^2 = \int_{-\infty}^{\infty} dt \psi_\beta(t)^2 \quad \text{and} \quad \|\chi\|^2 = \int_{-\infty}^{\infty} dt \chi(t)^2.$$

Furthermore, equality is attained in (12) if and only if the two functions are proportional:

$$|\langle \psi_\beta, \chi \rangle| = \|\psi_\beta\| \|\chi\| \iff \chi(t) = k \psi_\beta(t) \quad \text{for some constant } k. \quad (13)$$

Thus we might try to estimate the actual target trajectory by maximizing the absolute value of $\tilde{\chi}[\beta]$. However, the norm $\|\psi_\beta\|$ depends on β in general, so we cannot just look for a global maximum of $|\tilde{\chi}[\beta]|$. Instead, we can maximize $|\langle \psi_\beta, \chi \rangle| / \|\psi_\beta\|$. A preferable method is to define a new functional

$$\mathcal{E}[\beta] = 1 - \frac{|\langle \psi_\beta, \chi \rangle|}{\|\psi_\beta\| \|\chi\|}, \quad (14)$$

which we call the error functional. Then (12) and (13) state that

$$0 \leq \mathcal{E}[\beta] \leq 1 \quad \text{and} \quad \mathcal{E}[\beta] = 0 \iff \chi(t) = k \psi_\beta(t). \quad (15)$$

The error functional is nonnegative, and it vanishes if and only if the "trial return" $\psi_\beta(t)$ is indistinguishable (up to a constant factor) from the measured return $\chi(t)$.

Thus, to estimate the target trajectory, we need to minimize $\mathcal{E}[\beta]$. But two qualifications must be added here.

(A) We have not actually given a method for finding β so as to minimize $\mathcal{E}[\beta]$. This is a variational problem which must be analyzed separately. One strategy is to compute $\mathcal{E}[\beta]$ in parallel for a set of "likely suspects" β and then declare the trajectory with the least error to be the best estimate in the set. Furthermore, the value $\mathcal{E}[\beta]$ can be viewed as an estimate for the error, giving a rough idea of how "close" β is to the actual trajectory. For example, the trajectory of an accelerating target might be parameterized as

$$\beta(t) = (\mathbf{r}_0 + \mathbf{v}_0 t + \frac{1}{2} \mathbf{a} t^2, t),$$

where \mathbf{r}_0 , \mathbf{v}_0 and \mathbf{a} represent the unknown initial position, initial velocity, and acceleration (assumed uniform) of the target. Then $\mathcal{E}[\beta]$ reduces to an ordinary function $\mathcal{E}(\mathbf{r}_0, \mathbf{v}_0, \mathbf{a})$ of the target parameters, which must now be minimized. Of course, if the target's acceleration is actually nonuniform, then none of the above trajectories may have zero error. The one with the least error simply gives the best quadratic approximation.

(B) Even if we find a trajectory β for which $\mathcal{E}[\beta] = 0$, this does not guarantee that $\beta(t)$ is the actual target trajectory since $\mathcal{E}[\beta]$ is not one-to-one in general. In other words, the return may not determine the target trajectory uniquely. But clearly we can ask for no more than to find trial

trajectories whose returns $\psi_\beta(t)$ are indistinguishable from the measured return. To the extent that $\mathcal{E}[\beta]$ is not one-to-one, the radar/sonar problem is therefore *ambiguous*. The extent of that ambiguity in the present general context needs to be studied.

3. Relation to Wideband Ambiguity Functions

We now show that our "ambiguity functional" $\tilde{\chi}[\beta]$ is a generalization of the usual radar ambiguity functions. In most of the literature (for example, [10-13]), ambiguity functions depend on time and frequency. These are *windowed Fourier transforms* of the return and they apply only in *narrowband* situations. A more recent class of ambiguity functions depend on time and scale [5,14-20]. These are *wavelet transforms* of the return, and they apply to *wideband* as well as narrowband situations. But in both cases, the target is assumed to move at a uniform velocity. Hence neither the wideband nor the narrowband ambiguity functions do easily yield information about acceleration. By making a series of restrictive assumptions about the motions of the transmitter, the receiver, and the target, we will see that $\tilde{\chi}[\beta]$ indeed reduces to the usual wideband ambiguity function. Moreover, if the outgoing signal $\psi(t)$ is narrowband and the target does not move too rapidly, then the wideband ambiguity function further reduces to the narrowband ambiguity function. This establishes our claim that $\tilde{\chi}[\beta]$ is a far-reaching generalization of the concept of ambiguity functions.

The assumptions leading to the first reduction are as follows:

A. Monostatic radar:

Assume that the transmitter and the receiver are one and the same, and that both are at rest at the origin in space. That is,

$$\alpha(t) = \gamma(t) = (0, t). \quad (16)$$

B. Uniform target motion:

Assume that the target moves at a constant velocity \mathbf{v} , so that $\beta(t)$ has the form

$$\beta(t) = (\mathbf{r}_0 + \mathbf{v}t, t). \quad (17)$$

Our objective is to find the initial position \mathbf{r}_0 and the velocity \mathbf{v} . But in practice, targets are *tracked* with a narrow-beam antenna, so the *direction* of \mathbf{r}_0 (i.e., the unit vector $\mathbf{n} = \mathbf{r}_0/|\mathbf{r}_0|$) and the transversal ("crossrange") component of \mathbf{v} can be estimated separately from the antenna motion. We therefore need only to find the *range* $r_0 = |\mathbf{r}_0|$ and the radial velocity v , also called the *range rate*. Effectively, we can therefore make the following assumption.

C. No transversal motion (tracking mode):

The target moves directly toward or away from the radar site, i.e.,

$$\mathbf{v} = v\mathbf{n}, \quad \text{where} \quad \mathbf{n} = \frac{\mathbf{r}_0}{|\mathbf{r}_0|} = \frac{\mathbf{r}_0}{r_0}.$$

Thus

$$\beta(t) = (r(t)\mathbf{n}, t), \quad \text{where} \quad r(t) = r_0 + vt. \quad (18)$$

Under assumptions A - C, our objective is reduced to estimating the range r_0 and the range rate v . When these assumptions are introduced into Equation (9), a computation (performed in the Appendix) gives the simple result

$$\psi_\beta(t) = A(t)\psi(\sigma t - \tau),$$

(19)

where

$$\begin{aligned} A(t) &= \frac{g_\gamma g_\beta (1 + v/c)}{16\pi^2 r(t)^2} = \text{attenuation factor} \\ \sigma &= \frac{c - v}{c + v} = \text{time scaling factor} \\ \tau &= \frac{2r_0}{c + v} = \text{time shift.} \end{aligned} \quad (20)$$

Each quantity in (20) has a simple intuitive explanation:

1. Equation (19) states that the *instantaneous power* $\psi_\beta(t)^2$ of the return is inversely proportional to $r(t)^4$, in agreement with the *radar range equation* [11-13]. Hence the attenuation factor $A(t)$ is reasonable.
2. The scale factor σ represents the *Doppler effect*. If $0 < v < c$, then $0 < \sigma < 1$ and (19) states that the return is a *stretched* version of $\psi(t)$. If $-c < v < 0$, then $\sigma > 1$ and the return is a *compressed* version of $\psi(t)$. (The relation to the usual Doppler frequency shift is discussed below.)
3. Writing (19) as $\psi_\beta(t) = A(t)\psi(\sigma(t - \tau'))$, the quantity $\tau' = \tau/\sigma = 2r_0/(c - v)$ is interpreted as the *time delay* at $t = 0$ since the signal must travel a round-trip distance of $2r_0$ at the relative velocity $c - v$.

Equation (19) shows that under the assumptions A - C, our ambiguity functional reduces to an ordinary function of the scale factor σ and the time shift τ :

$$\tilde{X}[\beta] \rightarrow \tilde{X}(\sigma, \tau) \equiv \int_{-\infty}^{\infty} dt A(t) \psi(\sigma t - \tau) \chi(t). \quad (21)$$

In practice it can usually be assumed that $A(t)$ is approximately constant over the time interval where $\psi(\sigma t - \tau) \neq 0$, since $\psi(t)$ is time-limited (usually a pulse or a pulse train) and the range $r(t)$ does not change significantly while the reflection takes place ($r_0 \gg |vt|$ for all relevant values of t). In that case, (19) reduces to

$$\psi_\beta(t) \rightarrow A \psi_{\sigma, \tau}(t), \quad (22)$$

where

$$A = \frac{g_\gamma g_\beta (1 + v/c)}{16\pi^2 r_0^2}$$

and $\psi_{\sigma, \tau}(t)$ is defined by

$$\psi_{\sigma, \tau}(t) = \psi(\sigma t - \tau). \quad (23)$$

This two-parameter family of functions, labeled by the scale factor and the time shift, is called the *wavelet family* generated by $\psi(t)$.

Note: In most of the wavelet literature, including [5], it is customary to define the wavelets by

$$\psi_{s, \tau'}(t) = s^{-1/2} \psi\left(\frac{t - \tau'}{s}\right). \quad (24)$$

These are related to (23) by setting $s = 1/\sigma$ and $\tau' = \tau/\sigma$. The factor $s^{-1/2}$ in (24) is actually unnecessary, since its only purpose is to ensure that all the wavelets have the same energy, i.e.,

$\|\psi_{s,\tau'}\| = \|\psi\|$ for all values of (s, τ') . The entire theory works equally well without it. (Its absence is made up for later, in the reconstruction formula; see [5], pages 63 and 68.) With our normalization, a factor of $\sqrt{\sigma}$ will show up instead in Equation (26) below. We use the form (23) because it is simpler and more directly related to the radar problem. While the parameter s in (24) can be interpreted as the *time scale* (small s means fine scale and large s means coarse scale), our parameter σ can be interpreted as the *time resolution*: Large σ means high resolution, and small σ means low resolution.

With (22), the ambiguity functional reduces to

$$\tilde{\chi}(\sigma, \tau) = A \langle \psi_{\sigma, \tau}, \chi \rangle = A \int_{-\infty}^{\infty} dt \psi(\sigma t - \tau) \chi(t), \quad (25)$$

which is (apart from the factor A) the *wavelet transform* of the return χ with respect to the wavelet family $\psi_{\sigma, \tau}$. This is known as the *wideband ambiguity function*; see [5, 14–20]. We have thereby established a connection between the ambiguity functional $\tilde{\chi}[\beta]$ and some of the recent radar literature.

The wideband ambiguity function $\tilde{\chi}(\sigma, \tau)$ is used to estimate the range r_0 and range rate v (which characterize our *restricted* trajectory $\beta(t)$ in (18)) in exactly the same way as the general ambiguity functional $\tilde{\chi}[\beta]$ was used to estimate a *general* trajectory $\beta(t)$. Again Schwarz's inequality implies that

$$|\tilde{\chi}(\sigma, \tau)| \leq A \|\psi_{\sigma, \tau}\| \|\chi\|,$$

and equality holds if and only if $\chi(t) = k \psi_{\sigma, \tau}(t)$ for some constant k . But

$$\|\psi_{\sigma, \tau}\|^2 = \int_{-\infty}^{\infty} dt \psi(\sigma t - \tau)^2 = \sigma^{-1} \|\psi\|^2.$$

Therefore, to find the correct values of (σ, τ) , we can maximize $\sqrt{\sigma} |\tilde{\chi}(\sigma, \tau)|/A$ (noting that A depends on σ through v). Equivalently, we can minimize the *wideband error function*, defined by

$$\mathcal{E}(\sigma, \tau) = 1 - \frac{|\langle \psi_{\sigma, \tau}, \chi \rangle|}{\|\psi_{\sigma, \tau}\| \|\chi\|} = 1 - \frac{\sqrt{\sigma} |\tilde{\chi}(\sigma, \tau)|}{A \|\psi\| \|\chi\|}, \quad (26)$$

which satisfies

$$0 \leq \mathcal{E}(\sigma, \tau) \leq 1 \quad \text{and} \quad \mathcal{E}(\sigma, \tau) = 0 \iff \chi(t) = k \psi_{\sigma, \tau}(t).$$

$\mathcal{E}(\sigma, \tau)$ estimates the error in the wideband parameters (σ, τ) which characterize the trajectory (18). Once a good estimate is found, the range and range rate are determined by solving (20), which gives

$$r_0 = \frac{c\tau}{1+\sigma} \quad \text{and} \quad v = \left(\frac{1-\sigma}{1+\sigma} \right) c. \quad (27)$$

We have seen above that a uniform motion of the target results in a *scaling* of the return. But the common treatment of the Doppler effect in most of the literature is as a *uniform frequency shift*. The connection between these two views is made by the *narrowband approximation*, which assumes that the outgoing signal $\psi(t)$ is nearly harmonic in that its spectrum is concentrated in a narrow frequency band centered around a high carrier frequency, and the target velocity is small compared to the propagation speed c of the waves. (The latter assumption may break down in sonar, where

c is the speed of sound.) The narrowband approximation is treated in many references and will therefore not be discussed here. See, for example, [5,8,11,12,17,18].

4. Conclusions

Physical wavelets have been shown to generalize ordinary ("mathematical") wavelets, such as those given in (23) and (24). The physical wavelets originate in real-world problems, and they reduce to ordinary wavelets under simple conditions (monostatic radar, uniform target motion in the line of sight). The idea of ambiguity functions has been generalized to the context of such wavelets. The resulting scheme can be used to analyze general (accelerating or nonlinear) target motions, as well as general and independent motions by the transmitting and receiving antennas. However, the transmitters, targets, and receivers considered here have all been assumed to be *point objects*. The next objective of this research is to examine physical wavelets with *extended sources* rather than point sources. Such wavelets can have some *directivity*, hence they provide more realistic models for actual antenna or array outputs. They can be used to probe extended objects rather than merely point objects, and it is hoped that this will provide a method not only for *tracking* but also for *imaging*.

5. Appendix: Derivation of Equation (19)

To prove (19), recall that

$$\alpha(t) = \gamma(t) = (\mathbf{0}, t) \quad \text{and} \quad \beta(t) = ((r_0 + vt)\mathbf{n}, t) = (r(t)\mathbf{n}, t).$$

To simplify the notation we also define

$$\rho(t) = \frac{r(t)}{c} = \frac{r_0 + vt}{c}.$$

Denoting the wavelet received at $\beta(t)$ (before re-emission) by $\psi'(t)$, we have

$$\begin{aligned} \psi'(t) &\equiv (R_\beta E_\alpha \psi)(t) = g_\beta \int dt' G(\beta(t) - \alpha(t')) \psi(t') \\ &= g_\beta \int dt' G(r(t)\mathbf{n}, t - t') \psi(t') \\ &= g_\beta \int dt' \frac{\delta(t - t' - \rho(t))}{4\pi r(t)} \psi(t') \\ &= g_\beta \frac{\psi(t - \rho(t))}{4\pi r(t)}. \end{aligned} \tag{28}$$

Therefore the reflected wavelet observed back at $\gamma(t) = \alpha(t)$ is

$$\begin{aligned} \psi_\beta(t) &= g_\gamma g_\beta \int dt' G(\alpha(t) - \beta(t')) \psi'(t') \\ &= g_\gamma g_\beta \int dt' G(-r(t')\mathbf{n}, t - t') \psi'(t') \\ &= g_\gamma g_\beta \int dt' \frac{\delta(t - t' - \rho(t'))}{4\pi r(t')} \cdot \frac{\psi(t' - \rho(t'))}{4\pi r(t')}. \end{aligned} \tag{29}$$

But

$$\begin{aligned} t - t' - \rho(t') &= \frac{c(t - t') - r_0 - vt'}{c} = \frac{ct - r_0 - (c + v)t'}{c} \\ &= -\frac{c + v}{c} \left[t' - \frac{ct - r_0}{c + v} \right]. \end{aligned} \quad (30)$$

Using the shorthand

$$T = \frac{ct - r_0}{c + v},$$

(30) implies that

$$\delta(t - t' - \rho(t')) = \frac{c}{c + v} \delta(t' - T).$$

Therefore (29) gives

$$\psi_\beta(t) = g_\gamma g_\beta \frac{c}{c + v} \cdot \frac{\psi(T - \rho(T))}{[4\pi r(T)]^2}. \quad (31)$$

But

$$r(T) = r_0 + v \left(\frac{ct - r_0}{c + v} \right) = \frac{c}{c + v} r(t),$$

hence

$$\begin{aligned} T - \rho(T) &= T - \frac{r(T)}{c} = T - \frac{r(t)}{c + v} = \frac{ct - r_0}{c + v} - \frac{r_0 + vt}{c + v} \\ &= \left(\frac{c - v}{c + v} \right) t - \frac{2r_0}{c + v}. \end{aligned}$$

Thus (30) becomes

$$\psi_\beta(t) = \frac{g_\gamma g_\beta (1 + v/c)}{16\pi^2 r(t)^2} \psi \left(\left(\frac{c - v}{c + v} \right) t - \frac{2r_0}{c + v} \right),$$

which is indeed (19).

6. References

1. G. Kaiser, Space-time-scale analysis of electromagnetic waves, in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Victoria, Canada, 1992.
2. G. Kaiser and R.F. Streater, Windowed Radon transforms, analytic signals and the wave equation, in C.K. Chui, ed., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, New York, 1992, pp. 399-441.
3. G. Kaiser, Wavelet electrodynamics, *Physics Letters A* **168**(1992), 28-34.
4. G. Kaiser, Wavelet electrodynamics, Part II: Atomic composition of electromagnetic waves, *Applied and Computational Harmonic Analysis* **1**(1994), 246-260.
5. G. Kaiser, *A Friendly Guide to Wavelets*, Birkhäuser, Boston, 1994.
6. G. Kaiser, Remote sensing with electromagnetic and acoustic wavelets: A variational approach, in H. Szu, ed., *Wavelet Applications for Dual Use*, SPIE Conference Proceedings #2491, Orlando, FL, April, 1995.
7. G. Kaiser, A variational approach to radar, in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, October, 1995.

8. G. Kaiser, Physical wavelets and radar, to appear in the *IEEE Antennas and Propagation Magazine* (December 1995 or February 1996 issue).
9. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
10. P.M. Woodward, *Probability and Information Theory, with Applications to Radar*, Pergamon Press, London, 1953.
11. C.E. Cook and M. Bernfeld, *Radar Signals*, Academic Press, New York, 1967; republished by Artech House, Norwood, MA, 1993.
12. A.W. Rihaczek, *Principles of High-Resolution Radar*, McGraw-Hill, New York, 1968.
13. D.K. Barton,, *Modern Radar System Analysis*, Artech House, Norwood, MA, USA, 1988.
14. D.A. Swick, An ambiguity function independent of assumption about bandwidth and carrier frequency, *NRL Report #6471*, Washington, DC, 1966.
15. D.A. Swick, A review of wide-band ambiguity functions, *NRL Report #6994*, Washington, DC, 1969.
16. H. Naparst, Dense target signal processing, *IEEE Transactions on Information Theory* **37**(1991), 317-327.
17. L. Auslander and I. Gertner, Wideband ambiguity functions and the $a \cdot x + b$ group, in L. Auslander, T. Kailath, and S. Mitter, eds., *Signal Processing: Part I – Signal Processing Theory*, Springer-Verlag, New York, , 1990, pp. 1-12.
18. W. Miller, Topics in harmonic analysis with applications to radar and sonar, in R.E. Bluhart, W. Miller, and C.H. Wilcox, eds., *Radar and Sonar, Part I*, Springer-Verlag, New York, 1991.
19. P. Maas, Wideband approximation and wavelet transform, in F.A. Grünbaum, M. Bernfeld, and R.E. Bluhart, eds., *Radar and Sonar, Part II*, Springer-Verlag, New York, 1992.
20. S. Sowelam and A. Tewfik, Multiple class adaptive wideband radar target imaging, in H. Szu, ed., *Wavelet Applications for Dual Use*, SPIE Conference Proceedings #2491, Orlando, FL, April, 1995.

THE SYNTHESIS AND CHEMISTRY OF PEROXONITRITES AND
PEROXONITROUS ACID

Albert D. Kowalak
Professor
Department of Chemistry

University of Massachusetts Lowell
One University Ave.
Lowell, MA 01854

Final Report for;
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

University of Massachusetts Lowell

December 1995

THE SYNTHESIS AND CHEMISTRY OF PEROXONITRITES AND
PEROXONITROUS ACID

Albert D. Kowalak
Professor
Department of Chemistry
University of Lowell

Abstract

Peroxonitrous acid(HOONO) and its conjugate base salts, peroxonitrites, are being studied to understand the chemistry of these species which may be found in the atmosphere. Nitrogen containing compounds that have been burned release oxides into the atmosphere which produce nitric acid. Peroxonitrous acid may be an intermediate in the formation of nitric acid; however, this intermediate is very unstable and decomposes quickly. Reactions of H_2O_2 and NO_2^- at high pH and low temperatures are being studied to produce peroxonitrites. Further reactions have been done using acetic acid as solvent at room temperature. UV-Vis and FTIR spectroscopy are used to monitor the reaction and characterize the products. The regions of interest are from 850 to 1100 cm^{-1} in the IR and 300nm in the UV region. Electrochemical techniques are also being studied to synthesize and monitor reaction conditions.

THE SYNTHESIS AND CHEMISTRY OF PEROXONITRITES AND PEROXONITROUS ACID

Albert D. Kowalak

Introduction:

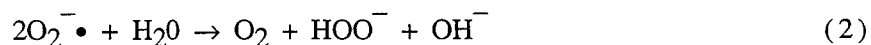
The chemistry of peroxonitrous acid and its conjugate base salts, peroxonitrites is an area of chemistry that needs study. Peroxonitrous acid, HOONO , has been observed as a reaction intermediate in the liquid phase, but has not been detected in the gas phase to date.¹ It has been suggested that peroxonitrites are present in soil samples studied in the Mars Viking experiments.² Preliminary kinetic studies of the reaction of the NO_2 radical with other radicals indicate the formation of two isomers of nitric acid, HONO_2 and HOONO .³ There is strong evidence that peroxonitrite is present in blood vessels.⁴ Nitrate is known to be present in smog droplets and in polar stratospheric clouds. The photoisomerization of nitrate in those changes that occur in the atmosphere is possible.

Nitrogen compounds when burned release nitrogen oxides into the atmosphere. These compounds react with moisture in the atmosphere to form nitric acid. This is one of the components of acid rain. Peroxonitrous acid can be considered as an intermediate in the formation of nitric acid. The peroxonitrous acid is very unstable and quickly decomposes into HNO_3 .

The presence of peroxonitrite in blood vessels from the reaction of nitric oxide and superoxide ion has been reported.



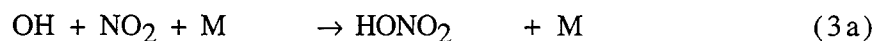
It was recently discovered that superoxide, $\text{O}_2^{\bullet -}$, in the body can be linked to the formation of peroxonitrite and a double-barreled toxicity can result.⁴ The radical $\text{ON}\bullet$ was found to be an important blood vessel dilating compound. With the addition of superoxide dimutase the lifetime of the dilating compound almost doubled. The fast reaction of $\text{ON}\bullet$ with $\text{O}_2^{\bullet -}$ in the above reaction forms peroxonitrite.⁵ It was concluded that the dilating compound was destroyed by the superoxide radical. The destruction of $\text{ON}\bullet$ was decreased if superoxide was disproportionate by the dimutase enzyme prior to the reaction of the two radicals.



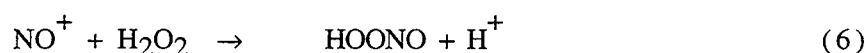
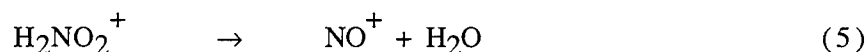
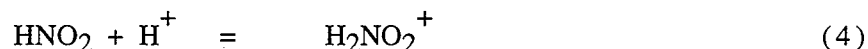
The two radicals react to form peroxonitrite, this anion is known to generate $\text{HO}\bullet$ under physiological conditions. Metabolic formation of $\text{ON}\bullet$ radical from the amino acid arginine, $\text{C}_6\text{H}_{14}\text{N}_4\text{O}_2$, has also reported in other sites where $\text{O}_2^{\bullet -}$ may be present. In both of these cases the peroxonitrite and hydroxyl radical should be formed and lead to tissue damage. The formation of ONOO^- in reaction (1) provides a path for the formation of peroxonitrous acid which can then decompose into nitric acid.

Regardless of the formation neither of these species are favorable in human blood vessels.

Detailed laboratory studies are needed to understand and investigate the chemistry and kinetics of these species that may be present in the atmosphere. Measurements of the stratosphere showed $\text{NO}_3^- (\text{HNO}_3)_n$ as one main series of negative ions present.⁶ These ions are generated from nitric acid and its isomer peroxonitrous acid according to reaction (3)



Previous studies of this reaction by Baeyer and Villiger show a preparation of peroxonitrous acid by mixing HNO_2 and H_2O_2 . Papée reported that the peroxonitrite ion was also produced by UV radiation of nitrate solutions with light of wavelength shorter than 280nm.⁷ The mechanism for the HNO_2 and H_2O_2 reaction has been proposed by Anbar and Taube⁸ to be:



Water and hydrogen peroxide compete for the NO^+ , nitrosonium cation, to form the peroxonitrous acid. In order to obtain ONOO^- the reactants must be at 0°C and strong base must be added immediately after the strong acid is added to initialize the reaction. Loss of peroxonitrous acid is due to isomerization of HOONO to nitric acid which has a time scale of 5s at 0°C and .2s at 27°C , by the following reaction.



Decomposition also contributes to the loss of peroxonitrous acid.



A preliminary rate law for the reaction of nitrous acid with hydrogen peroxide has been reported to be:

$$-d[\text{HNO}_2]/dt = k [\text{HNO}_2][\text{H}_2\text{O}_2][\text{H}^+] \quad (9)$$

Stabilizing peroxonitrites by this procedure is difficult because at low pH the isomerization of HOONO is faster than its formation.

First Method of Synthesis:

Our first method of synthesis involved the formation of peroxonitrite using hydrogen peroxide and nitrite. Many reactions were done by varying the amounts of the two reactants. These results can be seen in table 1 in the appendix on page A1. After these reactions were conducted a .025 M NaNO₂ solution and a 2% H₂O₂ solution along with the resultant mixture of the two solutions were analyzed using UV-VIS spectroscopy. The percent transmittance is given at three wavelengths, and can be seen in table 2 in the appendix on page A1.

After reviewing these results we decided to synthesize peroxonitrite by reacting equal amounts of nitrite, peroxide, base, and acid. A one molar sodium nitrite solution was prepared in one molar sodium hydroxide. This was done by

dissolving 6.9 grams of dry sodium nitrite in one molar sodium hydroxide. The hydrogen peroxide was prepared by diluting 30% peroxide with distilled water. One molar perchloric acid was used as the proton source, and one molar sodium hydroxide was used as the strong base. These starting materials were chilled to 0°C in an ice bath and kept in separate containers. Mixing of the starting materials is done in a certain way. First NaNO₂ and NaOH are added together. Then this mixture and the peroxide are added to the perchloric acid simultaneously. The resultant mixture is the final product and it is placed in an ice bath at 0°C. A yellow solution was obtained and believed to be ONOO⁻. This product had a pH of approximately 12.0 and was analyzed using UV-VIS and FTIR spectroscopy.

Analysis of the product using UV-VIS spectroscopy shows a strong absorption at 300nm. This agrees with the previously reported color and intense absorption at 302nm with an $\epsilon_{\text{max}} = 1670 \pm 50 \text{ M}^{-1} \text{ cm}^{-1}$.⁹ This technique is difficult to use because the intense color distorts the absorption, and when the solution is diluted in order to be analyzed it warms up and decomposes. When decomposition takes place the solution becomes clear. Analysis by FTIR spectroscopy using silver bromide cells shows peaks in the 850 to 1100 cm⁻¹ region. This is the region of spectra that should contain peroxide linkages, see spectra in appendix on page B1. The peak at 1110 cm⁻¹ disappears over time as the solution begins to decomposes. After decomposition has taken place, the peak at 1110 cm⁻¹ has completely disappeared. Analysis of the starting products show a lack of any significant peaks in this region, see spectra B2,3,4,5. This indicates that the peroxide linkage is present and decomposes over time as proposed. The intense yellow color also disappears over time which further leads us to believe that the product synthesized is peroxonitrite. Previous research done separately by McGrath and Burkholder show peaks in this region.²⁶

Second Method of Synthesis:

A second method of synthesis was considered and several reactions with acetic acid as the proton source were performed. The amounts of nitrite, peroxide, and acetic acid were varied and the results can be seen in Table 2 on page A2. After several reactions were run the specific amounts of the starting materials were discovered. It was found that glacial acetic acid, one molar sodium nitrite dissolved in distilled water, and 6% peroxide were used as starting materials. The previous condition of 0°C can not be met using glacial acetic acid because it freezes. Equal quantities of the three starting materials are used to produce the peroxonitrous acid. All three of the starting materials are placed in a temperature controlled water bath at 20°C in separate containers. The peroxide and nitrite are added simultaneously to the acid. This gives a yellow solution that liberates heat, and is presumed to be peroxonitrous acid.

Analysis of the yellow product using UV-VIS spectroscopy shows absorption in the specified region.

The starting materials were also analyzed using UV-VIS spectroscopy and no peaks were seen in the region of 300nm. Preliminary analysis using FTIR shows peaks in the specified region, see spectra B6. This would seem to indicate that the final product is HOONO and it is acidic. One problem with this result is that glacial acetic acid also has peaks in the same region that peroxide linkages are observed. These peaks interfere with the peaks of peroxonitrous acid, see spectra B7. This method of synthesis is very important because it seems to lead to the desired product; however, complete analysis of the product has not yet been done to confirm the formation of peroxonitrous acid.

Third Method of Synthesis:

A third method of synthesis was attempted using phosphoric acid as the proton source. Several preliminary reactions were done using different concentrations of phosphoric acid with peroxide and nitrite. The phosphoric acid concentration was varied from pure to .1M with peroxide concentrations of 3 and 6%. A one molar sodium nitrite solution was used as the nitrite source. Several different conditions were used including those used for the previous synthetic techniques. The results of each reaction was a blue solution that liberated brown gas and heat.

This product is believed to be dissolved NO_2 gas and possibly an intermediate in the formation of N_2O_5 or N_2O_4 . Phosphoric acid seems to be too strong of an acid to use without the nitrite. Solid sodium nitrite was added to phosphoric acid and a blue solution with brown gas was produced. Because of these results this method of synthesis was not pursued further.

Other Techniques:

Other techniques were attempted to improve on the previous synthetic techniques. Manganese dioxide, MnO_2 , was added to several of the solutions of peroxonitrous acid as a peroxide scavenger. This created excessive amounts of oxygen and many bubbles that destroyed the peroxonitrous acid. The mixing and vigorous bubbling of the product added oxygen to the peroxonitrous acid and mixed the solution enough to decompose the acid. A reaction was also done using solid sodium nitrite. A small amount of solid sodium nitrite was placed in a beaker and

equal amounts of glacial acetic acid and 12% peroxide were added. A yellow solution with a slight amount of brown gas formed. Nitrogen gas was bubbled through the product to drive off the brown gas and a yellow solution resulted. This solution was capped and put on the shelf to be analyzed.

Electrochemistry:

Electrochemical techniques were also attempted to synthesize peroxonitrous acid. Electrochemistry is considered because of the possibility of an oxidation reduction reaction taking place to produce the desired product. The assigned oxidation state of nitrogen in HNO_3 is 5, and in HOONO is 3 due to the peroxide linkage. The oxidation state of nitrogen in NO_2^- is also 3. This would seem to indicate that no oxidation or reduction is being carried out in the conversion of NO_2^- to HOONO . However, this may not be the case because the OOH linkage definitely changes the molecule and may have an effect on the oxidation number of the nitrogen.

A palladium/hydrogen electrode was used as a reference with platinum loops as the working and counter electrodes. One molar perchloric acid was used as the electrolyte. A Pt/PtO cyclic was run to verify that the system was working properly. The CV is run using perchloric acid that has been bubbled with argon gas to remove any oxygen. If the cyclic is observed to be of the correct form, reactions are attempted. Several cycles were done using different concentrations of NaNO_2 and H_2O_2 . The yellow solution was never obtained but some results were of interest. This technique needs to be continued and studied further.

Conclusion:

The formation of peroxonitrite by the traditional methods has been accomplished and improved upon.

Synthesis of peroxonitrous acid using acetic acid as the proton source has been attempted and positive results have been obtained. Once this technique is improved the synthesis and analysis of the acid can be completed. Electrochemical techniques have also been attempted and notable results have been obtained.

Recommendations:

Continuation of this research should include the complete analysis of peroxonitrous acid in acetic acid, along with an improved synthesis technique that reduces the amount of acetic acid need to perform the reaction. If this can be accomplished the product can be analyzed using FTIR spectroscopy and only the peroxonitrous acid peaks will be in the region of interest. Reactions using peroxyacetic, peracetic acid, as the initializing acid should also be considered in continued research. Preliminary reactions have been conducted using a peroxidase enzyme in the synthesis of our product, this line of synthesis should also be explored further. Another possibility for the continuation of the research is the formation of metal compounds with peroxynitrites. Some preliminary reactions were performed using PbNO_3 and pernitrite. These reactions have not been carried out to completion, but plans for future research include reactions of this type.

Literature Cited:

1. Burkholder, J.B., Hammer, P.D., & Howard, C.J., *J. Phys. Chem.*, **91**, 2136 (1987).
2. Plumb, R.C., Tantayanon, R., Libby, M., & Xu, W.W., *Nature (London)*, 338, 633 (1989).
3. Niki, H., Maker, P.D., Savage, C.M., & Breitenbach, L.P., *Chem. Phys. Lett.*, **1978**, 59, 78.
4. Beckman, J.S., Beckman, T.W., Chen, J., Marshall, P.A., & Freeman, B.A., *Proc. Natl. Acad. Sci. USA*, **87**, 620 (1990).
5. Blough N.V. and Zafiriou, O.G., *Inorg. Chem.*, **24**, 3502 (1985).
6. Viggiano, A.A., *Mass Spectrom. Rev.*, **1993**, 12, 115.
7. Petriconi, G.L. and Papée, H.M., *Can J. Chem.*, **44**, 977 (1966).
8. Anbar, M., & Taube, H., *J. Amer. Chem. Soc.*, **76**, 6243 (1954).
9. Hughes, M.N. and Nicklin, H.G., *J. Chem. Soc. A*, 3179 (1968).

Kevin Malloy report unavailable at time of publication.

THEORETICAL STUDIES OF THE PERFORMANCE OF A NOVEL FIBER-COUPLED
IMAGING INTERFEROMETER

Sudhakar Prasad
Associate Professor of Physics and Astronomy

University of New Mexico
800 Yale Blvd. NE
Albuquerque, NM 87131

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

University of New Mexico
Albuquerque, NM

March 12, 1997

THEORETICAL STUDIES OF THE PERFORMANCE OF A NOVEL FIBER-COUPLED IMAGING INTERFEROMETER

Sudhakar Prasad
Associate Professor of Physics and Astronomy
University of New Mexico

ABSTRACT

This document describes the work done in accomplishing the technical objectives of Contract No. F49620-93-C-0063. The three broad objectives originally proposed were partially accomplished in the course of the work. The problem of computing the fluctuations of the coupling efficiency of the multiple-core (MC) fiber as well as the problem of computation of the theoretical sensitivity of an N -telescope interferometer based on MC fibers were entirely solved. The problem of computer simulations of the performance of such interferometers has also been partially solved. Rather unexpectedly, it turned out that the atmospheric simulation code that we used was not entirely adequate for our purposes, and so an extensive program of developing a satisfactory code has now been undertaken. Because of this unexpected development, the problem of the simulation of image reconstruction using standard deconvolution algorithms could not be undertaken except in a rather rudimentary way.

PREFACE

This document describes the work done in accomplishing the technical objectives of Contract No. F49620-93-C-0063. Sec. 1 provides an introduction to the subject matter of the report. In Sections 2 through 4, the work done to meet the main tasks of the Contract are summarized. Sections 2 and 3 have an associated Appendix that contains a publication covering in detail the research accomplishing the main task of the Contract. In Sec. 2, discussions and results of our study of the coupling efficiency and its fluctuations in MC fibers are presented, while in Sec. 3 we present some details of the theoretical calculations of the sensitivity with which fringe power and bispectra can be measured by our interferometer. In Sec. 4, we describe the simulations that we have performed on MC-fiber-coupled interferometer under conditions of atmospheric turbulence but without photon noise.

It is a pleasure to acknowledge the technical and research support of Air Force PL/LIMI at all stages of this Contract. I am particularly thankful to Gary Loos and Sergio Restaino, who have at all stages of this work been generous with their time and always expressed enthusiastic support for the work I have performed. Concomitant support for the work described here has also come under a Contract with the Air Force PL/LIMI, for which I am grateful.

CONTENTS

<u>Section</u>		<u>Page</u>
1.0	INTRODUCTION	5
2.0	COUPLING FLUCTUATIONS	7
2.1	MEAN COUPLING EFFICIENCY	8
2.2	VARIANCE OF COUPLING EFFICIENCY	9
3.0	MC-FIBER-COUPLED INTERFEROMETER:	
	THEORETICAL RESULTS	11
3.1	MEAN, VARIANCE, SNR OF $ \mu_{12} ^2$	11
3.2	MEAN, VARIANCE, SNR OF BISPECTRUM	12
4.0	COMPUTER SIMULATIONS OF FRINGE POWER, BISPECTRA	14
4.1	ATMOSPHERIC SIMULATION	14
4.2	FIBER SIMULATION	16
	REFERENCES	19

1.0. INTRODUCTION

A large variety of applications of single-mode (SM) optical fibers arise from their unique properties. They include mechanical flexibility and strength, perfect spatial coherence of the guided mode, the realization of ultra-low linear propagation losses as well as novel nonlinear optical pulse propagation and shaping effects. Recently, a new application area for SM fibers has emerged. In a typical optical telescope, light from an astronomical source undergoes a large number of reflections (typically 20 or so), with a high attendant power loss, before being incident on the imaging detector. One can imagine coupling starlight into SM fibers (Ref. 1) early in the beam train, for once coupled light can be transported in an essentially lossless manner via the spatially coherent guided mode for processing and detection. It is furthermore easy to combine and split fiber beams via fiber couplers. All of these features are especially relevant for a ground-based high-angular-resolution interferometric array of telescopes (Ref. 2), in which an image is reconstructed from information gathered in the spatial frequency domain via telescope-to-telescope field correlations whose measurement involves beam splitting and recombination. Additionally, the perfect spatial coherence of coupled light implies high fringe visibilities that are limited solely by the source structure and almost completely independent of atmospheric or telescope seeing conditions (Ref. 3).

In spite of all these advantages of SM fibers, they suffer from one serious drawback, namely their woefully small coupling of degraded wavefronts under conditions of atmospheric turbulence. A new approach, proposed by the author (Refs. 4 and 5), is to employ a multiple-core fiber concept with enhanced coupling efficiencies when exposed to phase-distorted wavefronts. A multiple-core fiber contains several cores immersed in a single cladding material in a highly symmetrical arrangement. The presence of multiple cores predisposes the fiber to an enhanced coupling efficiency because the distorted wavefront has to only piecewise match the SM excitations of each core. After the coupling of a wavefront into such a fiber at the focal plane of a telescope, the multiple cores can then be physically separated into individual single-core SM fibers, each carrying a sub-beam, for the purposes of interferometric beam combination with similar sub-beams from other telescopes of the imaging array. In this way, one accomplishes both highly efficient coupling as well as beam splitting, the two integral aspects of interferometry.

To demonstrate that fibers are of any use in interferometry, one has to analyze the sensitivity of fiber-coupled ground-based arrays with respect to image reconstruction, and show that the sensitivity is at the very least not compromised by the use of MC fibers. We present results on the work we have done with regard to the image of a point source for

which the sensitivity is a maximum, and the best-case performance of the interferometer is attained. The interferometer sensitivity is limited at high-light levels by the fluctuations of the coupling efficiency and by the intercore correlations of the MC fibers. A study of such fluctuations is presented in Sec. 2. Both theoretical calculations and computer simulations based on atmospheric phase-screens obeying Kolmogoroff statistics were performed. The results of this part of the work are presented in Secs. 3 and 4.

2.0. MC-FIBER COUPLING FLUCTUATIONS

Here we merely summarize the useful results; the details can be found in the Appendix which contains Ref. 6.

The efficiency with which a wavefront can be coupled into a single-mode fiber depends directly on the overlap of the wavefront with the spatial mode profile. For light of frequency ck and telescope of focal length f , the coupling efficiency takes the form

$$\eta = \sum_{\lambda} \eta_{\lambda}, \quad (1)$$

where the partial couplig efficiency for the λ th mode of the MC fiber is

$$\eta_{\lambda} \equiv \frac{4\rho_0^2 k^2}{\pi^2 D^2 f^2} \int' \int' \tilde{E}_0(\vec{\rho}) \tilde{E}_0^*(\vec{\rho}') e^{-\frac{k^2}{2f^2}(\rho^2 + \rho'^2)\rho_0^2} \frac{S_{\lambda}(\vec{\rho}) S_{\lambda}^*(\vec{\rho}')}{(N_{\lambda}/\pi\rho_0^2)} d^2\vec{\rho} d^2\vec{\rho}'. \quad (2)$$

The mode-dependent function $S_{\lambda}(\vec{\rho})$ has the expression

$$S_{\lambda}(\vec{\rho}) = \sum_{j=0}^{N-1} \alpha_j^{(\lambda)*} e^{-i(k/f)\vec{p}_j \cdot \vec{\rho}}, \quad (3)$$

in which the α 's specify the complex excitation amplitudes of the individual cores for a particular fiber mode.

When the incident wave is phase-corrupted by atmospheric turbulence, the normalized pupil field $\tilde{E}_0(\vec{\rho}) = \exp[i\phi(\vec{\rho})]$ becomes a statistical quantity with a Gaussian statistical phase ϕ with the following bilinear moment:

$$\gamma_{\mu}(|\vec{\rho} - \vec{\rho}'|) \equiv \langle \tilde{E}_0(\vec{\rho}) \tilde{E}_0^*(\vec{\rho}') \rangle = e^{-3.44(|\vec{\rho} - \vec{\rho}'|/r_0)^{5/3} [1 - \mu(|\vec{\rho} - \vec{\rho}'|/D)^{1/3}]}, \quad (4)$$

where $\mu = 0$ for a fully uncorrected phasefront and 1 when its global tilt is removed frame by frame.

For the purposes of computing and interpreting the SNRs of squared visibility, bispectra, and reconstructed image, it is useful to derive simple formulas that are valid in the asymptotic regime of large pupil diameters, $D \gg r_0$. In this limit, the image field is highly speckled with field correlations only within each speckle. It is expected that for an N -core MC fiber, no more than about one speckle per core, i.e., N speckles in all, may be coupled into the fiber. This would imply a power coupling efficiency of about $(r_0/D)^2$ for a SCSM fiber and N times that for an N -core MC fiber. We make this physical picture quantitatively precise by considering the average and variance of the exact expression (2) for coupling efficiency in this asymptotic limit.

2.1. MEAN COUPLING EFFICIENCY

The average value of the coupling efficiency is given by replacing inside (2) the product of normalized pupil fields by its average (4). With this replacement, it is clear that when the double pupil integral is performed, only points $\vec{\rho}$ and $\vec{\rho}'$ that are no more than $O(r_0)$ away will contribute significantly. This is made more precise by transforming the $\vec{\rho}'$ integral to be over the difference vector $\vec{R} = \vec{\rho}' - \vec{\rho}$, so that only R less than $O(r_0)$ will contribute. It may be seen by explicit algebra that since the optimal ρ_0 is about $\pi f/(kD)$ (see Ref. 4) and since $r_0 \ll D$, one may replace $S_\lambda(\vec{\rho}')$ by $S_\lambda(\vec{\rho})$ and $\exp[-(k^2/2f^2)\rho'^2]$ by $\exp[-(k^2/2f^2)\rho^2]$ inside the average of integral (3). This procedure will decouple the $\vec{\rho}$ integration from the \vec{R} integration, the latter of which can be extended over the whole infinite plane in the same limit $r_0 \ll D$. In this way, the average coupling efficiency may be seen to be well approximated by the following three lowest order terms in r_0/D :

$$\begin{aligned} \langle \eta \rangle &\approx \frac{4\rho_0^2 k^2}{\pi^2 D^2 f^2} \int d^2 \vec{R} \gamma_0(R) \left[1 + 3.44\mu(R/r_0)^2 \left(\frac{r_0}{D}\right)^{1/3} + \frac{1}{2} 3.44^2 \mu^2 (R/r_0)^4 \left(\frac{r_0}{D}\right)^{2/3} \right] \\ &\quad \times \int' d^2 \vec{\rho} e^{-\frac{k^2}{f^2} \rho^2 \rho_0^2} \sum_\lambda \frac{|S_\lambda(\vec{\rho})|^2}{(N_\lambda/\pi\rho^2)} \\ &\approx \left(\frac{r_0}{D}\right)^2 \left[1 + 1.06\mu \left(\frac{r_0}{D}\right)^{1/3} + 1.24\mu^2 \left(\frac{r_0}{D}\right)^{2/3} \left(1 - e^{-k^2 \rho_0^2 D^2 / 4f^2}\right) (F_N/F_1) \right]. \end{aligned} \quad (5)$$

The factor (F_N/F_1) defines the enhancement of average coupling efficiency for an N -core MC fiber over the standard single-core fiber, with F_N given by

$$F_N \equiv \sum_\lambda \left[\int' d^2 \vec{\rho} e^{-k^2 \rho_0^2 \rho^2 / f^2} |S_\lambda(\vec{\rho})|^2 \right] / (N_\lambda/\pi\rho_0^2). \quad (6)$$

It must be kept in mind that the above result is not valid for ρ_0 much larger than $\pi f/kD$, which for maximum coupling is roughly its optimal value alluded to earlier. In fact, the coupling should slowly decrease with increasing ρ_0 with which increases the number of mutually incoherent speckles intercepted by the fiber that tend to cancel one another's contribution to the mode overlap integrals that determine the coupling efficiency.

The difference from the single-core single-mode (SCSM) fiber with regard to mean coupling is merely due to the factor F_N/F_1 . This factor may be easily evaluated for sufficiently large ρ_0 . When $\rho_0 > \pi f/kD$, the exponential factor inside the expression (6) for F_N falls to a very small value at the outer limit of radial integration, namely at $D/2$. This means that limit may be extended to ∞ without much error. With this simplification, the integration in (6) may be carried out, when expression (3) for $S_\lambda(\vec{\rho})$ is

substituted. For each λ , when the rotational symmetry of the core centers is exploited, the N^2 terms obtained, one for each term in the double sum resulting from the use of (3) inside (6), become F_1 times $N_\lambda/(\pi\rho_0^2)$, the latter canceling with an identical term in the denominator. Thus each λ term on the right hand side (RHS) of (6) is just F_1 for sufficiently large ρ_0 . Since there are as many modes as cores, this result implies

$$F_N/F_1 = N. \quad (7)$$

Thus, when $D/r_0 \gg 1$, the maximum mean coupling efficiency for an N -core fiber is just N times that for an identical single-core fiber, exactly what we expect based on the independent speckle picture alluded to earlier.

2.2. VARIANCE OF COUPLING EFFICIENCY

As a measure of the fluctuation of the coupling efficiency, we consider its variance for a general MC fiber. The variance σ_η^2 ,

$$\sigma_\eta^2 = \langle \eta^2 \rangle - \langle \eta \rangle^2, \quad (8)$$

may be expressed, in view of (2), in terms of mode-mode covariances as

$$\sigma_\eta^2 = \sum_\lambda \sum_{\lambda'} [\langle \eta_\lambda \eta_{\lambda'} \rangle - \langle \eta_\lambda \rangle \langle \eta_{\lambda'} \rangle]. \quad (9)$$

Firstly, for D/r_0 , $\langle \eta_\lambda \rangle$ is independent of λ , as seen at the end of Sec. 2.1, and thus equal to the SCSM mean coupling efficiency. Secondly, with the expression (2) for η_λ , it is easy to see that $\langle \eta_\lambda \eta_{\lambda'} \rangle$ involves a quartic pupil integral with integrand involving the average of a product of 4 pupil field values. This quadrilinear field average can be evaluated in terms of the phase structure function under the assumption of Gaussian statistics for the field phase, but the quadruple pupil integral is numerically too daunting to compute accurately. We shall instead adopt an approximate procedure valid in the asymptotic regime of large D/r_0 .

For large D/r_0 , one may decompose each pupil into many - roughly $(D/r_0)^2$ - mutually independent coherence cells, each of diameter r_0 , so that each pupil integral is in effect a sum of a large number of independent random variables, and as such has Gaussian statistics. Equivalently, in our quadruple integral, the pupil field itself may be taken as a Gaussian variable. This enables us to write for $\langle \eta_\lambda \eta_{\lambda'} \rangle$ the following expression:

$$\langle \eta_\lambda \eta_{\lambda'} \rangle = \langle \eta_\lambda \rangle \langle \eta_{\lambda'} \rangle + |g_{\lambda\lambda'}|^2, \quad (10)$$

where one may show that

$$g_{\lambda\lambda'} = \delta_{\lambda\lambda'} \langle \eta_{\lambda} \rangle. \quad (11)$$

Thus, one obtains for the coupling efficiency variance:

$$\sigma_{\eta}^2 = \langle \eta \rangle^2 / N. \quad (12)$$

This result represents the physical understanding that the variance of a sum of N independent random variables is just the sum of individual variances, all equal in our case. Consequently, the ratio of the standard deviation of the sum to its mean (also called relative dispersion) is $1/\sqrt{N}$ that for each individual term of the sum. With N cores, the relative fluctuation of overall coupling is reduced, since a speckle missed by one core has a finite chance still of being coupled into the other $N - 1$ cores.

3.0. MC-FIBER-COUPLED INTERFEROMETER: THEORETICAL RESULTS

Several important features of MC fiber coupling that we have described in the previous section, namely high coupling efficiency, ease of beam splitting, high core-core correlations under evanescent mixing, and all other features that accrue to any *coherent* coupling scheme (spatial filtering, insensitivity to atmospheric turbulence, etc.) have enabled us to propose an MC-fiber-coupled interferometer (Ref. 6). A schematic diagram of the interferometer is displayed in Fig. 1.

Since the phase of the normalized complex fringe visibility is often impossible to measure directly, interferometers typically measure two other quantities, namely the fringe visibility modulus squared (also known as fringe power spectrum) and fringe triple products or bispectra. The latter are defined as the ordered product of 3 complex fringe visibilities around a closed triangle of 3 apertures. Thus, the bispectrum b_{123} is the product $\mu_{12}\mu_{23}\mu_{31}$, which is independent of aperture dependent (piston) phases characteristic of atmospheric turbulence while has a phase that is the sum of the true complex visibility phases, namely $\phi_{12} + \phi_{23} + \phi_{31}$. If enough triple sum phases, also known as closure phases, are measured and standard a priori constraints used, then one may recover the individual fringe phases and therefore the object intensity structure without corruption by the atmosphere.

The sensitivities with which the fringe power and bispectra may be measured are dependent on the total number of photons captured by the individual apertures as well as on the coupling efficiency fluctuations induced by atmospheric turbulence. Theoretical calculations of these sensitivities were carried out in great detail in Ref. 6; here we shall only reproduce essential results from that paper.

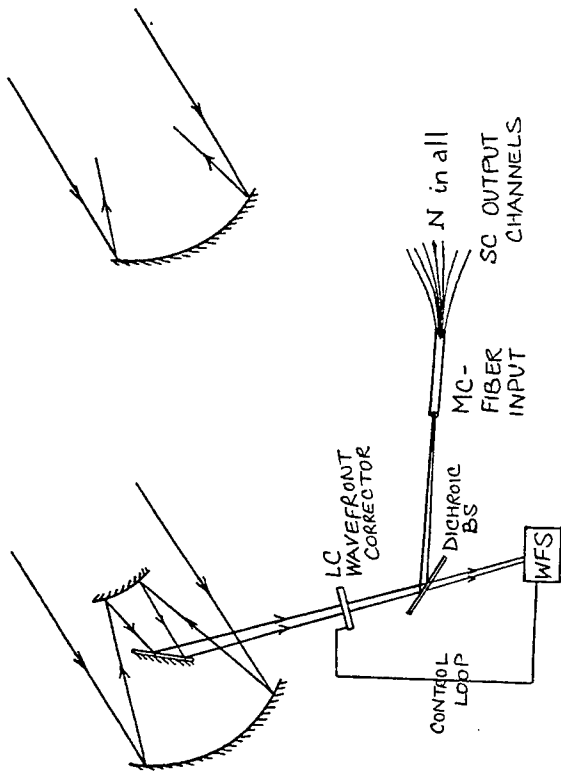
3.1. MEAN, VARIANCE, AND SNR OF MEASUREMENT OF $|\mu_{12}|^2$

The correlation modulus squared, $|\mu_{12}|^2$, may be estimated by measuring $|z_{12}|^2$ given by

$$|z_{12}|^2 = (N_+ - N_-)^2 + (N'_+ - N'_-)^2, \quad (13)$$

where N_{\pm}, N'_{\pm} represent photon counts registered in the \pm output arms without and with relative phase shift as indicated in Fig. 1. After correcting $|z_{12}|^2$ for a certain well-understood photon bias, one may compute its average and variance for arbitrary photon numbers. After straightforward steps, the SNR characterizing the measurement of fringe power turns out to be

Schematic of A MC-Fiber-Coupled Interferometer



- N telescopes, each using an N -core fiber
 $\implies N^2$ in all

* 1 for Fringe Tracking $+(N-1)$ for Pairwise Correlation

- Pair-wise beam combination via difference detection

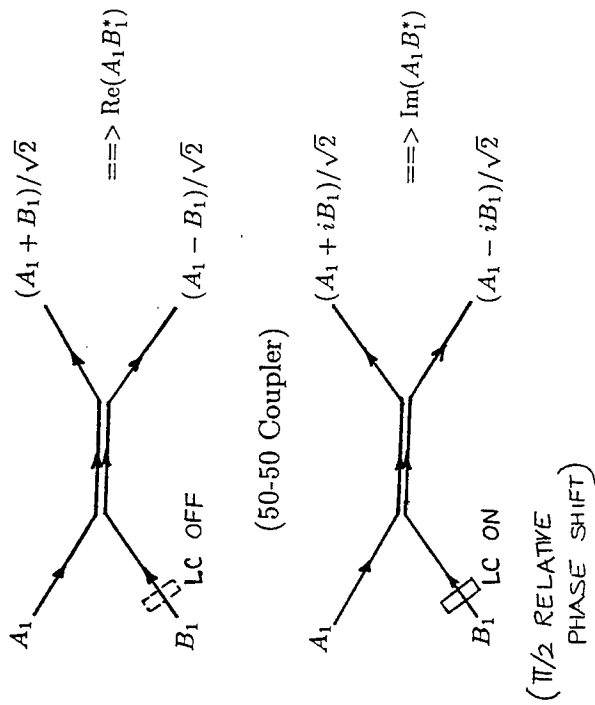


Fig. 1: A schematic diagram of a MC-fiber-coupled interferometer. Only one element is shown explicitly. The 50-50 fiber coupler proposed for pair-wise beam recombination is also displayed along with the manner in which correlation information is to be obtained.

$$\text{SNR}(|Z_{12}|^2) = \frac{|\mu_{12}|^2 \langle \eta_1 \rangle^2 N_0^2}{\sqrt{(2\chi_1^2 + \chi_1^4)|\mu_{12}|^4 \langle \eta_1 \rangle^4 N_0^4 + 2(1 + \chi_1^2) \langle \eta_1 \rangle^3 N_0^3 |\mu_{12}|^2 + (1 + 0.5\chi_1^2) \langle \eta_1 \rangle^2 N_0^2}} \quad (14)$$

where χ_1^2 is the fractional fluctuation of coupling efficiency:

$$\chi_1^2 \equiv \frac{\text{var}(\eta_1)}{\langle \eta_1 \rangle^2} \quad (15)$$

and $\langle \eta_1 \rangle$ is the average coupling fraction into a single strand of each MC fiber.

Let us now compare the performance of MC-fiber-based array to a single-core fiber-based array with respect to fringe power measurements. In the MC-fiber setup, the fractional fluctuation of coupling efficiency changes as a function of the distance of propagation in the MC fiber before the individual strands are separated at some optimal distance $z = L$ (described in Ref. 5). In fact, due to evanescent coupling of the field inside the various cores that maximizes core-core correlation at appropriate distances, it is expected that the fractional fluctuation χ_1^2 will be *smaller* for the MC fiber than when a SC fiber is used with N -fold beam splitting, particularly when D/r_0 is not too big, less than about 10 (see Sec. 8 for simulation results). (For larger D/r_0 , evanescent coupling is ineffective in correlating fields, since mutually incoherent speckles are coupled into the cores.) This clearly implies a higher SNR $(2\chi_1^2 + \chi_1^4)^{-1/2}$ at high light levels. At low light levels too, a MC-fiber-based system should have higher SNR because of a larger $\langle \eta_1 \rangle$, e.g., between 3 (for $D/r_0 = 5$) and 7 ($D/r_0 > 20$) times the equivalent figure for SC-fiber-based system.

3.2. MEAN, VARIANCE, AND SNR OF MEASUREMENT OF BISPECTRUM b_{123}

The calculations for the SNR of a bispectrum are more involved, since the bispectrum contains a triple product of fringe phasors rather than the modulus squared of a fringe phasor. A principal difference between the two cases is a sharp reduction of the bispectrum signal under conditions of moderate to high atmospheric turbulence ($D/r_0 > 5$), while the noise behavior is little affected.

For complex quantities like the bispectrum, the square root of its pseudo-variance, which is the following positive number:

$$V(B_{123}) = \overline{|B_{123}|^2} - |\overline{B_{123}}|^2, \quad (15)$$

is a better representative of the size of the nearly circular noise cloud (for which the true variance is nearly zero) that surrounds the mean value on a phasor diagram. With this as the noise, the following expression for the SNR of the measurement of bispectrum is obtained:

$$\begin{aligned} \text{SNR}(B_{123}) &\equiv \frac{|\overline{B_{123}}|}{\sqrt{V(B_{123})}} \\ &= \frac{\gamma^3 \langle \eta_1 \rangle^3 N_0^3 \mu_{12} \mu_{23} \mu_{31}}{\sqrt{\langle \eta_1 \rangle^6 N_0^6 |\mu_{12}|^2 |\mu_{23}|^2 |\mu_{31}|^2 (1 + 3|\gamma|^2 + 3|\gamma|^4) + \langle \eta_1 \rangle^3 N_0^3 (1 + 0.75|\gamma|^2)}} \quad (16) \end{aligned}$$

where γ is approximately the core-core normalized correlation referred to above.

The preceding SNR takes the following expressions under low-light and high-light levels:

$$\lim_{N_0 \rightarrow 0} \text{SNR}(B_{123}) = \frac{|\gamma|^3 \langle \eta_1 \rangle^{3/2} N_0^{3/2} \mu_{12} \mu_{23} \mu_{31}}{\sqrt{(1 + 0.75|\gamma|^2)}}. \quad (17)$$

and

$$\lim_{N_0 \rightarrow \infty} \text{SNR}(B_{123}) = \frac{|\gamma|^3}{\sqrt{(1 + 3|\gamma|^2 + 3|\gamma|^4)}}. \quad (18)$$

At high photon numbers, the SNR of the bispectrum is degraded for the MC fiber based array, principally because of the overall γ^3 factor in the numerator of expression (18). This is because the main contribution to the pseudo-variance comes from auto-correlations of the baselines, while it is their weak cross-correlations that determine the bispectrum signal. At low photon numbers, however, the SNRs for the SC-fiber and MC-fiber systems are comparable, because of the improved coupling efficiency for the latter system.

For a value of about 0.5 for γ expected for a 7-core MC fiber operating in the intermediate $D/r_0 \approx 5$ regime, a single strand of the MC fiber is about 3 times more efficiently coupled than the equivalent figure for a SC fiber: $\langle \eta_1 \rangle_{MC} \approx 3\langle \eta_1 \rangle_{SC}$. With $\gamma = 1$ for a SC fiber, whose sub-beams are perfect clones of each other, this translates into an SNR for the 7-core fiber that is only 20% lower than for a SC fiber at low photon numbers, although the relative performance is a factor of 4 degraded at high photon numbers. It is of course performance at low photon numbers that determines the limiting magnitude of sources being imaged by an interferometer. Under such conditions, a MC-fiber-based system provides for a tremendous ease of beam splitting and recombination without much loss of sensitivity.

4.0. COMPUTER SIMULATIONS OF FRINGE POWER AND BISPECTRUM

Simulations have been performed to verify the enhancements in coupling efficiency and sensitivity that are predicted for a ground-based stellar interferometer array using multiple-core, single-mode (MCSM) optical fibers. The performance of such an instrument has been well characterized theoretically, assuming idealized Kolmogoroff statistics for the atmospheric fluctuations, and resulting phase aberrations, in the light from a point source incident on elements of the array. Simulations with an appropriate ensemble of computer-generated phase screens provide non-ideal fluctuations for a more realistic check of the theory. The simulation also includes complexities, such as differences in the core-core correlations within a MCSM fiber, that are not considered in the theoretical analysis. It will also be possible to investigate the effects of photon and detector noise, perturbation of core positions, and atmospheric outer scale with the simulation. The goal of the simulation was to compare the performance of MCSM fibers to conventional single-core fibers by calculating and comparing coupling efficiency and its standard deviation, as well as signal-to-noise ratios of the visibility modulus squared and the bispectrum. The program which performs the simulation can be divided into a part which generates the phase screens and a part which determines the fiber modes, the interaction with the incident field, the propagation through the fibers, and the recombining to give visibilities and bispectra.

4.1 ATMOSPHERIC SIMULATION

A truly realistic simulation of light passing through the turbulent atmosphere would include fluctuations in both amplitude and phase of the light. In practice, amplitude variations (scintillation) are usually ignored, and only phase fluctuations, which fundamentally limit the resolution (for a single telescope) and the fringe visibility (for an interferometer) are considered. A map of the phase amplitude $p(x, y)$ over an aperture defines the field, or phase screen: $\exp[-ip(x, y)]$.

There are two common approaches for creating phase screens. The first, proposed by McGlamery [7], consists of generating an array of normally distributed random numbers in the spatial frequency domain, multiplying the array with a filter function chosen to represent the turbulence statistics (obtained from the Wiener power spectrum of phase fluctuations), where the contribution of the piston phase is removed, and performing an inverse Fourier transform to obtain the spatial phase map. This approach suffers from a high frequency cutoff due to the failure to capture the phase fluctuations above the

Nyquist frequency, and a low frequency cutoff from underestimating the power in the low-order spatial modes. The high-frequency problem can be solved with a finer grid spacing, but simultaneously resolving the low-frequency problem would require excessively large matrices.

The second approach uses an expansion of Zernike polynomials to estimate the phase. When the phase over a circular aperture is expanded as a weighted summation of these polynomials, the covariance between the coefficients of the summation can be computed. The resulting covariance matrix is given by Roddier [8], based on work by Noll [9]. Because of the non-zero off-diagonal terms in this matrix, it is not possible to generate independent random amplitudes for each Zernike term in a simulation. It is necessary to diagonalize the covariance matrix: a similarity transformation to a space which is spanned by an orthogonal set of uncorrelated Karhunen-Loeve functions. The eigenvalues of the diagonalized matrix are the variances of these functions, and the eigenvectors of the transformation matrix are representations of the K-L functions in terms of Zernike polynomials. We can treat the K-L function amplitudes as uncorrelated normal random variables of given variance in the simulation, generate a set of random amplitudes, then get the correct Zernike weights for the phase summation through the transformation matrix of eigenvectors. This approach suffers at high spatial frequencies due to the numerical precision necessary to represent the high order terms.

The method used in this simulation is a hybrid approach developed by Cochran [10], which resolves some of the difficulties of the above methods. The algorithm generates a map of random phase fluctuations in the spatial frequency domain. The map is filtered and Fourier transformed to the spatial domain. Then a K-L expansion is performed to give a finite set of correctly weighted Zernike modes. Only the first few modes (low spatial frequencies) need to be represented in this way. These modes are projected out of the original phase map and added back in with the correct weighting.

4.2 FIBER SIMULATION

The collective electromagnetic modes, or “supermodes”, of the MCSM fiber core ensemble are determined using a variational approach, including the eigenvectors describing the relative excitations of the fiber cores for each mode, the associated eigenvalue propagation constants, and the optimal mode-width parameter for the summation of weighted Gaussian modes. The spatial mode distributions are calculated in the image plane, stored

in an array, and transformed to the pupil plane in order to calculate the overlap with the incident field. All of this needs to be done only once given fixed fiber parameters. For each frame and aperture, the overlap integral between the field and mode distribution is calculated, giving the coupling efficiency, and the core excitations are calculated as a function of propagation distance through the fiber. The fiber cores are separated at the point at which the core-core correlations are maximized, and complex correlations are determined for each distinct pair of cores from separate apertures. This information is used to calculate the visibility modulus squared, and bispectra from “triangles” of apertures. These quantities are averaged over frames to give the signal-to-noise ratios. The theoretical SNR for visibility is calculated from the fractional fluctuation of coupling efficiency for a single core at the point of separation, and the theoretical bispectrum SNR is found from the average core-core correlation (between cores within each of the MCSM fibers).

The simulation was done for a 7-element array, with each telescope pupil coupling light into a 7-core MCSM fiber. The source was a 632.8nm point source which produced phase corrupted plane waves over the apertures. Each telescope was assumed to have perfect tilt correction. The simulation neglected photon noise, corresponding to the high-light-level limit, where the only noise source is the fluctuation in fiber coupling efficiency. Statistics were calculated over ensembles consisting of 100 “frames”, or iterations representing a coherence time of the atmosphere. For each frame, a statistically independent phase screen was projected on each pupil of the 7-element array. Different ensembles were generated for ratios of pupil diameter D to Fried parameter r_0 of 1.0, 5.0 and 10.0. Telescopes with $D=0.5\text{m}$ and focal length $f=2.8\text{m}$ were used. The secondary mirror obscuration was ignored. The pupil pixel scale was 1.0cm and the array size was 128x128 pixels. The cores of the fibers were symmetrically placed in a hexagonal geometry inside the cladding material. The fibers had a V parameter of 2.0, and a ratio of core separation to radius of 3.0. The ratio of core radius to the diameter of the diffraction-limited Airy disk in the image plane was 0.5, corresponding approximately to the predicted value for maximum coupling efficiency.

4.3 RESULTS AND DISCUSSION

The results obtained from the phase-screen simulation of wavefronts and their subsequent coupling and processing by the MC-fiber-coupled interferometer are presented below and compared with the theoretical results of Refs. 5 and 6. All of the results have

to be interpreted with caution because of an unexplained variability of the phase-screen simulations from one set of 50-100 frames to the next set.

In our simulation, we generated all 21 complex fringe phasors that can be formed out of pair-wise combinations of sub-beams from our 7 apertures. In order to simulate the experimental situation, we use the first 6 cores (arbitrarily labeled as such) of each MC fiber, skipping the last core, and combine each such core beam with a core beam from each of the remaining 6 apertures. This mimics the experimental situation in which physically distinct fiber cross-couplers are employed for measuring different fringe phasors. Once the 21 fringe phasors are obtained in this way, then their moduli are squared and the 35 distinct, ordered triple products of those fringe phasors, i.e., the 35 bispectra are formed. Complete information about the phases of the latter, the so-called closure phases, and the fringe power spectra enables the reconstruction of the object intensity structure. We therefore simulated the mean, standard deviation and their ratio, the SNR, for both the power spectra and the bispectra to make a preliminary assessment of the performance of our interferometer. Complete simulation of the performance would require simulating certain objects and then reconstructing the actual intensity maps from the measured and apriori information. The latter was not done because of a lack of time.

The results of simulation of the SNR's of the 21 power spectra were quite comparable to the theoretical predicted values. To make comparisons meaningful, we first frame-averaged each of the 21 power spectra, also computing their SNRs over all of the frames, typically 50 in number, and then computed a grand average SNR and standard deviation from the 21 SNR averages computed above. This produced a single mean SNR which was then compared with the theoretical prediction. For $D/r_0 = 1, 5, 10$, the simulated mean SNRs were 3.07, 0.82, and 0.62, respectively, while the corresponding predicted numbers (based on the high-light formula presented in the last paragraph of Sec. 6.1) were 5.3, 0.66, and 0.47. The discrepancies in the above numbers between the simulated and theoretically predicted results have probably to do with the inaccuracies of wavefront simulation used for the purpose as well as the small numbers of frames used (see discussion in Ref. 9 on the subject).

The simulations for the bispectra on the other hand are quite different from the predicted values. A similar approach as for the fringe power spectra was used to condense the whole set of 35 bispectra and their frame-averaged SNRs into a single grand-average SNR number for each D/r_0 value. For $D/r_0 = 1, 5, 10$, the simulated mean SNRs for the bispectrum were 1.09, 0.08, and 0.17, respectively, while the corresponding predicted numbers (based on the high-light formula (18)) were 0.32, 0.02, and 0.002. The huge

discrepancies between the simulation and theoretical-prediction numbers are not at all understood at present. Although the reasons here must include those that plague a similar comparison for the power spectra described in the previous paragraph, the situation is exacerbated by the very sensitive dependence of the bispectrum signal on the core-core correlations (which have no effect on the fringe power).

In an attempt to resolve the above discrepancies, we have begun investigating the details of the phase-screen simulation package that we borrowed from C. Matson [11] to do the simulations described in this report. A preliminary examination of the simulated phase screens shows that the simulated phase structure function, which is an important measure of the accuracy of simulation, departs significantly from the Kolmogoroff-Fried prediction for it. However, since our theoretical predictions do not refer to a particular turbulence model in their most general form, rather they are based on parameters like core-core correlations and coupling efficiency fluctuations directly computed by the simulation, we cannot explain away the discrepancies between the simulation and prediction merely as a consequence of incorrect atmospheric modeling. More work needs to be done to understand these results.

REFERENCES

1. S. Shaklan and F. Roddier, "Coupling starlight into single-mode fiber optics," *Appl. Opt.* **27**, 2334-2338 (1987).
2. C. Froehly, "Scientific importance of high angular resolution at infrared and optical wavelengths," in *Proc. ESO Conf. High Angular Resolution*, M. Ulrich and K. Kj  r, eds. (ESO, Garching, Germany, 1981), pp. 285-293; S. Shaklan and F. Roddier, "Single mode fiber optics in a long-baseline interferometer," *Appl. Opt.* **26**, 2159-2163 (1987); V. Coud   du Foresto and S. Ridgeway, "FLUOR: a stellar interferometer using single-mode infrared fibers," in *Proc. High Resolution Imaging by Interferometry II*, J. Beckers and F. Merkle eds. (ESO, Garching, Germany, 1991), pp. 731-740.
3. P. Connes, S. Shaklan, and F. Roddier, "A fiber-linked ground-based array," in *Proc. Joint NOAO-ESO Workshop on High-Resolution Imaging from the Ground Using Interferometric Techniques*, Oracle, Arizona, J. Goad, ed. (NOAO, Tucson, Arizona, 1987), pp. 165-168.
4. S. Prasad, "Focusing light into a multiple-core fiber: I. Theory," *Opt. Commun.* **115**, 354 (1995).
5. S. Prasad, "Focusing light into a multiple-core fiber: II. Application to ground-based interferometry," *Opt. Commun.* **115**, 368 (1995).
6. S. Prasad, "Coupling and Sensitivity Enhancement with Multiple-Core Fibers in Ground-Based Interferometry," in *Fiber Optics in Astronomical Applications*, Samuel C. Barden, Editor, *Proc. SPIE* 2476, 96-107 (1995).
7. B.L. McGlamery, "Computer simulation studies of compensation of turbulence degraded images," *Proc. SPIE* **74**, 225 (1976).
8. N. Roddier, "Atmospheric wavefront simulation and Zernike polynomials," *Proc. SPIE* **1237**, 668 (1990).
9. R. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**, 207 (1976).
10. G. Cochran, "Phase screen generation," *The Optical Sciences Company* (1985).
11. C. Matson, private communication.

STATIC AND DYNAMIC GRAPH EMBEDDING FOR PARALLEL PROGRAMMING

Mark Purtill
Assistant Professor
Department of Mathematics

Texas A&M University-Kingsville
Campus Box 172
Kingsville TX 78363

Final Report for
Summer Research Extension Program
Phillips Laboratory

Sponsored by
Air Force Office of Scientific Research
Bolling AFB, Washington DC

and

Phillips Laboratory, Albuquerque NM

December 1995

STATIC AND DYNAMIC GRAPH EMBEDDING FOR PARALLEL PROGRAMMING

Mark Purtill
Assistant Professor
Department of Mathematics
Texas A&M University-Kingsville

Abstract

Many problems in computational physics can be viewed as computations carried out on graphs (networks); one such problem arises in particle-in-cell (PIC) plasma simulations. As many multiprocessors (massively parallel computers) can also be viewed as graphs, the problem of partitioning such problems to run efficiently on a multiprocessor is a graph theoretic problem. It can be viewed as choosing a graph map between the two graphs minimizing a function.

Unfortunately, this problem is NP-complete, which means it is very unlikely that we can find an efficient, exact algorithm. Therefore, we must rely on heuristic methods. Many such methods have been proposed for this and related problems.

To investigate which of these algorithms are best suited for the graphs arising from PIC simulations, we have developed a test bed program which runs on both networks of workstations and on multiprocessors. The test bed is written in C++, using modern programming methods such as object-oriented design and literate programming.

Mark Purtil

1 The Problem¹

1.1 Graph Theory Background

A *graph* G consists of a set of objects, called *vertices* (singular *vertex*), denoted $V(G)$ and a set of unordered pairs of vertices, called *edges*, denoted $E(G)$. (For our purposes, both of these sets must be finite.) The two vertices in an edge are called its *ends*, and the edge is said to *connect* its ends. We do not allow an edge to connect a vertex to itself, nor for more than one edge to connect a given pair of vertices (that is, our graphs are “simple”).

A graph may be represented by a drawing (picture), such as figure 1; however, the geometric information (such as coordinates of the dot representing a vertex) are *not* part of the graph; figure 2 represents the same graph G as figure 1: the vertex set is $V(G) = \{a, b, c, d\}$ and the edge set $E(G) = \{\{a, b\}, \{a, c\}, \{b, c\}, \{b, d\}, \{c, d\}\}$.

If every pair of vertices is an edge (that is, all pairs of vertices are connected), then the graph is called “complete”. Figure 3 is a complete graph on five vertices.

Note that in some cases (such as figure 3), it is necessary for the edges of the drawing to cross; this has no effect on the graph itself. Graphs that can be drawn without crossings are called planar graphs; these have various nice properties (see *e.g.* [8]). We do not assume our graphs are planar, since many interesting graphs are not planar.

A sequence of vertices $(v_1, v_2, v_3, \dots, v_n)$ such that each pair of vertices v_i, v_{i+1} is called a *path*. For instance, in the graph in figure 1, (a, c, b) is a path, while (a, d, b) is not. (Note that we can also specify a

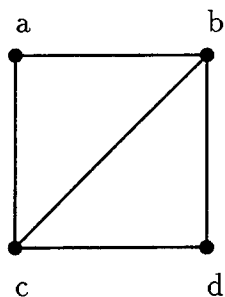


Figure 1: A drawing of a graph; the disks represent vertices, and the lines represent edges.

¹Much of the background material in this report is taken from my SFRP report for 1995.

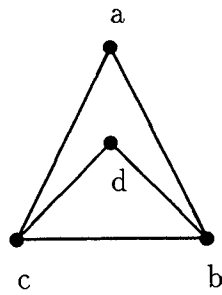


Figure 2: A different drawing of the same graph as in figure 1.

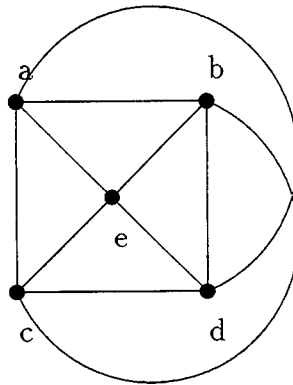


Figure 3: A drawing of a graph which cannot be drawn without edges crossing: the complete graph on five vertices.

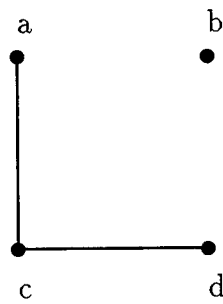


Figure 4: This graph is disconnected because (*e.g.*) there is no path from vertex *a* to vertex *b*.

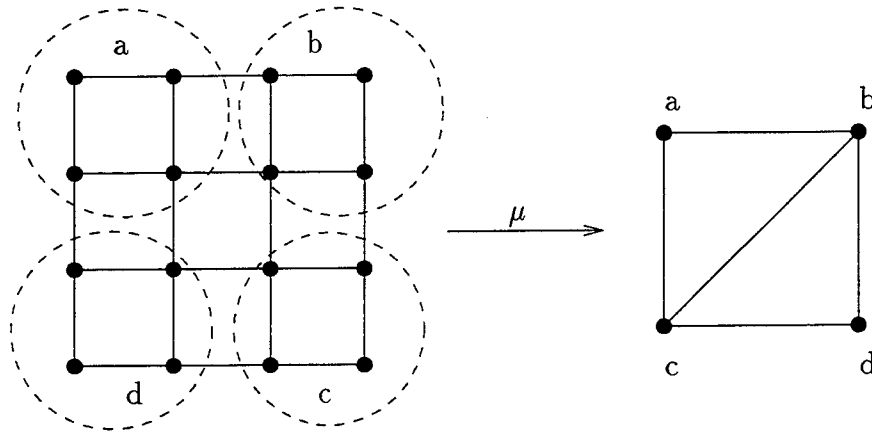


Figure 5: A graph map; each of the dashed circles in the graph on the left hold the vertices in the preimage of the vertex in the graph on the right with the same label.

path by listing the edges in that path rather than the vertices.) The *length* of a path is the number of *edges* in it.

If there is a path between any pair of vertices in the graph, then the graph is called *connected*. The graphs in figures 1, 2 and 3 are connected; that in figure 4 is not connected. All graphs considered in this paper will be connected. The *distance* between two vertices in a graph is the length of the shortest path connecting them; the *diameter* of the graph is the longest distance between two vertices. For instance, the graph in figure 1 has diameter 2.

A *graph map* μ from a graph G to a graph H , denoted $\mu: G \rightarrow H$ is a function μ from the vertex set of G , $V(G)$, to the vertex set of H , $V(H)$, so that for every edge $\{v, w\}$ of G , either $\mu(v) = \mu(w)$ or $\{\mu(v), \mu(w)\}$ is an edge of H ; that is:

$$\text{if } \{v, w\} \in V(G), \text{ then either } \{\mu(v), \mu(w)\} \in V(H) \text{ or } \mu(v) = \mu(w).$$

(Aside for graph theorists: if we assume each vertex of each graph is equipped with a self-loop, this coincides with the usual definition of *e.g.* [8].)

For each vertex v of H , we denote by $\mu^{-1}(v)$ the set of all vertices of G which map under μ to the vertex v (the *preimage* of v); that is $\mu^{-1}(v) := \{w \in V(G) \mid \mu(w) = v\}$. A sample graph map is shown in figure 5.

The *valance* or *degree* of a vertex is the number of edges of which it is an endpoint, or, equivalently, the number of vertices which are connected to it by an edge. In figure 1, vertices a and d have valance 2, which b and c have valance 3. If all of the vertices of a graph have the same valance, the graph is called *regular*; for instance, figure 3 shows a regular graph (of valance 4).

For a simple application of the graph concept, consider a multiprocessor; that is, computer consisting of a number of different processors, each with its own memory, which can communicate with each other *via*

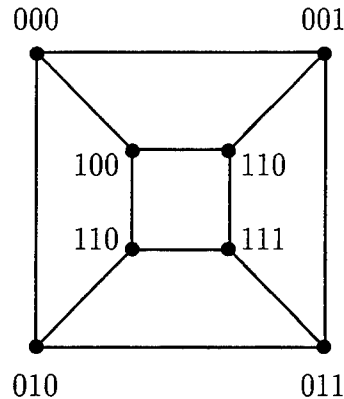


Figure 6: A three-dimensional hypercube graph.

communications links. Assume each link connects exactly two processors.

This arrangement of processors can be expressed as a graph; the vertex set of the graph is the set of processors, while the edge set is all the pairs of processors which are directly connected. (So the set of edges is the same as the set of links.) When we think of a graph as representing a multiprocessor, we will refer to the vertices (edges) as processors (links).

A path (p_1, p_2, \dots, p_n) in this graph is a way that a message could be passed from processor p_1 to processor p_n through processors p_2 through p_{n-1} via direct links, and the graph is connected precisely when each processor can communicate with each other processor (directly or indirectly).

For example, a common network connection is the “ n -dimensional hypercube”, which consists of 2^n processors, labeled 0 to $2^n - 1$. Two processors are connected if the binary representation of their labels differ in exactly one bit. See, for example, figure 6. For instance, the computer we did some work on this problem in 1994 (an Intel iPSC/860 called *neutrino*—see [21]) had the topology of the four-dimensional hypercube.

Often it is handy to be able to associate a number with each vertex or each edge of the graph; we call such numbers *weights*. Vertex weights are numbers associated with each vertex of the graph, while edge weights are numbers associated with each edge. For instance, we might have vertex weights giving the relative speed of each processor, and edge weights giving the speed or bandwidth of each communications link.

Formally, if we say a graph G has “vertex weights w ”, we mean w is a function from the vertex set $V(G)$ of G to the real numbers, and write $w(v)$ for the weight of vertex v . Similarly, if we say G has “edge weights u ”, we mean u is a function from the edge set $E(G)$ of G to the real numbers. In this paper, we will only need vertex and edge weights which have non-negative values.

1.2 Motivation from Plasma Physics Simulation

Many problems involving numeric solving of partial differential equations can be cast as graph computation. Here we describe one such problem, the particle-in-cell method for numerical computation of medium density

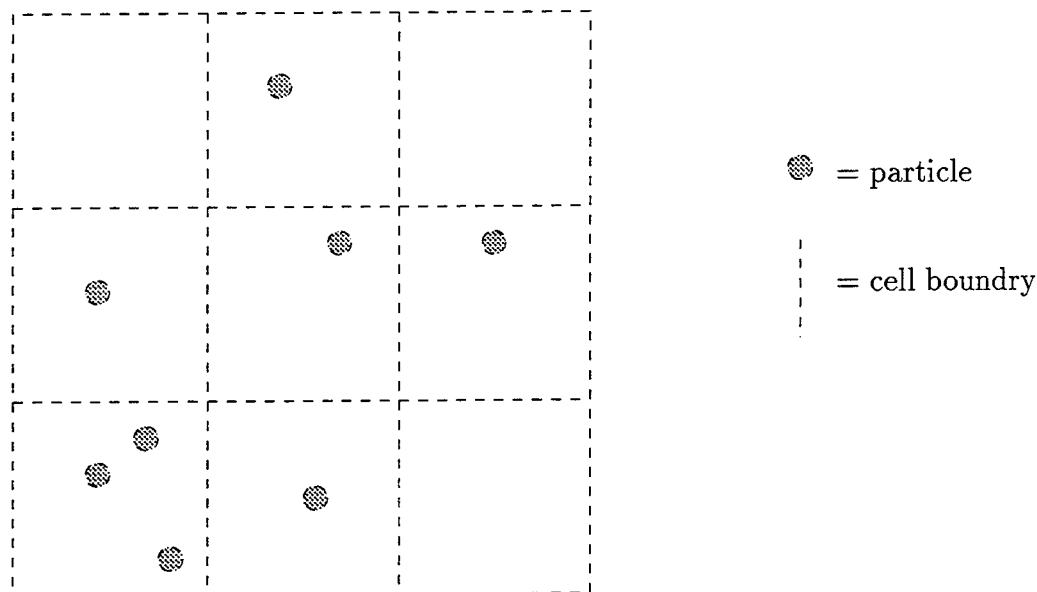


Figure 7: Particle-in-cell plasma simulation.

plasma (PIC plasma simulation).

In a plasma, there are magnetic and electrical fields interacting with charged particles moving in space; the interaction is governed by Maxwell's equations (for the fields) and the Lorentz equations (for the motions of the particles). In the particle-in-cell model, space is divided into cubes (cells) by a grid. The fields are kept track of on the grid points (or, in more sophisticated codes, on faces or edges of the cubes, with different field components having different locations). The particles, on the other hand, are allowed to move freely; they are kept track of by which cell they are in, then the coordinates of the particle within the cell. For more details, see (for instance) Birdsall and Langdon [2].

We can construct a graph to represent this computation. The vertex set of the graph is the set of cells of the spatial decomposition, while the set of edges is the pairs of cells which must exchange information in order for the computation to proceed. This information exchange occurs two ways in this problem:

- particles sometimes will move between adjacent cells; and
- adjacent cells must exchange field information in order to "step" Maxwell's equations.

Depending on the exact details, and whether it is necessary to keep track of both kinds of communication, the exact details of the edges may vary; see figure 8 for a typical example. (In this case, we assume that particle movement between diagonally connected cells is sufficiently rare that transferring the particle *via* another cell is acceptable. Another option would be to use edge weights.)

The computation graphs we get from PIC codes will have a nice structure; for instance, "most" cells (all those not on the boundary) will have the same valance; thus, the graph is "close" to being regular. In the case of figure 8, for instance, the center cell is "typical" (in the sense that a finer subdivision of space would give

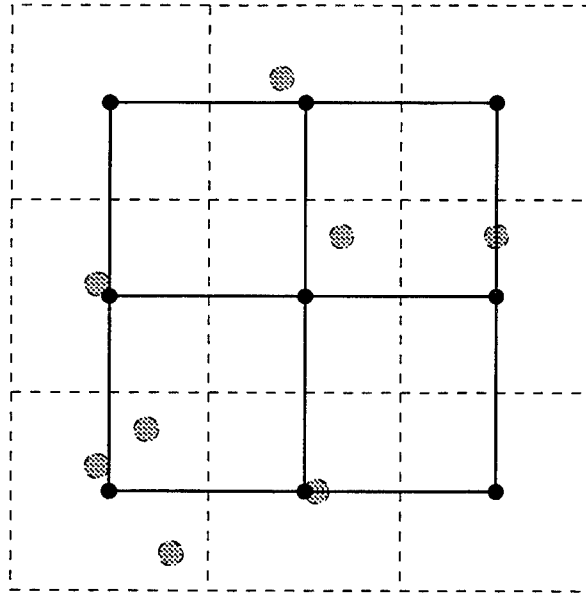


Figure 8: A two-dimensional subdivision of space (dashed), together with a corresponding graph. Note that the even distribution of space does not result in an even balance of particles.

a large number of cells which look like that cell); its valance is 4. (It is not clear whether this near-regularity would be useful.)

Because of the movement of particles, PIC simulations have an additional feature compared with some other problems: the amount of work to be done at the vertices representing the cells will change over time as particles move in the system. The amount of computation needed will be (in general) will be unpredictable. (At a given time step, amount of computation needed could be given by vertex weights.)

Now, suppose we wish to run a PIC plasma simulation on a multiprocessor. For maximum efficiency, we would like to assign cells to different processors in such a way that adjacent cells map to either the same processor or processors connected by a communications link, in such a way that processors have (as close as possible to) equivalent load and not too much communications is necessary.

This can be reformulated in the language of graph theory: the assignment of cells to processors is exactly a graph map μ from the PIC plasma simulation's graph G to the multiprocessors graph M ; this is the statement that adjacent cells map to the same or connected processors. If the processors and links of the multiprocessor are all the same, and the vertices and edges of the PIC plasma simulation graph have equal load on them, then the statement that the processors are equally loaded reduces to requiring that the sizes of the $\mu^{-1}(v)$ be the same, and the requirement that there be a small amount of communications becomes the requirement that there be a small number of edges connecting the $\mu^{-1}(v)$ s. In case of differing speeds of processors and so on, vertex and edge weights can be used to capture the differences.

The problem we are mainly concerned with is a generalization of the problem just discussed; we assume

that there is already a map μ , which is not necessarily optimal, and we wish to find a new map ν which is load balanced, has low communication, and is close to μ . If we think of the map μ as an assignment of vertices of the PIC graph to processors of a multiprocessor, then this last desire says that not too many vertices must be moved between processors to change to the new assignments. The relative weightings of the three desires would depend on a number of factors; for instance, if the problem is near completion, we would not be willing to spend a lot of time to speed up the last few steps, and thus the last desire would become more important; at the beginning of a long run, it would be less important or even ignored entirely.

1.3 Mathematical Formulation

We can formulate this “Graph Remapping” problem in mathematical terms. Using techniques of [6], we can reduce it even the “mapping” problem to an “NP-complete” problem; this means it is extremely unlikely that there is an algorithm which solves the mapping (or remapping) problem optimally in time polynomial in the size of the input.

Consider two graphs, G and M .

The graph G , representing the problem to be solved, has vertex weights w and edge weights c : for a vertex $v \in V(G)$, $w(v)$ is the amount of work needed to be done at v during one step of the computation; for an edge $e \in E(G)$, $c(e)$ is the amount of communications needed between the two ends of the edge. Borrowing from graph embedding, we will call G the “guest graph”.

For the graph M , representing the multiprocessor the problem will be solved on, we call the vertices *processors* and the edges *links*. The graph M has vertex weights s and edge weights b : for a processor $p \in V(M)$, $s(p)$ is the speed of processor p ; for a link $l \in E(M)$, $b(l)$ is the bandwidth of the link l . Again, borrowing from graph embedding, we will call M the “host graph”².

Given a graph map $\mu: G \rightarrow M$, the problem is to find a new graph map $\nu: G \rightarrow M$ minimizing some increasing function of

- $move_v = \sum_{v \in V(G)} d_M(\mu(v), \nu(v))$ for each $v \in V(G)$;
- $work_p = \frac{1}{s(p)} \sum_{v \in \nu^{-1}(p)} w(v)$ for each $p \in V(M)$; and
- $comm_e = \frac{1}{b(l)} \sum_{e \in \nu^{-1}(l)} c(e)$ for each $l \in E(M)$.

For instance, the function f given by

$$f(\nu) = \max(\alpha move_v) + \max(\{\beta work_p \mid p \in V(M)\} \cup \{\gamma comm_e \mid e \in E(M)\}),$$

²This is also appropriate because M represents a multiprocessor “host”.

with suitable choices of the constants α , β , and γ , could measure the total time needed to load balance (first term) and run a step (second term) on a multiprocessor which can overlap commutation and communications.³

We call the general problem "Graph Remapping"⁴; we assume that the map μ represents a solution to the same problem at a previous time step. If the function to minimize takes no account of the previous map μ , then this reduces to the problem of "Graph Mapping"; this could be used to initially load the multiprocessor; alternatively, depending on the multiprocessors I/O capabilities, we might wish to start with a map μ which maps all the vertices of G to a single processor of M .

Because of the motivation from parallel computing, we are most interested in solving this problem on the multiprocessor modeled by graph M . Since communications is only possible on edges of the graph, any algorithm which balances the load requires at least the diameter of M communications steps. (Recall that the diameter of a graph is the longest distance between two vertices, or, in this case, processors. If few communications steps than the diameter are taken, then two processors which are the diameter apart cannot communicate, and thus cannot average loads with on another.)

1.4 NP-completeness

This algorithm can be considered as a decision problem by giving a bound F on $f(\nu)$ and asking if any mapping ν exists for which $F \geq f(\nu)$. In this form, the theory of NP-completeness can be applied to the problem. In the format of the standard reference in that area, Garey and Johnson [6], the problem would be expressed as follows (with our sample f given above):

MINIMAL GRAPH REMAPPING

INSTANCE: Graphs G and M and a graph map $\mu: G \rightarrow M$. Four nonnegative numbers K , α , β , and γ .

QUESTION: Is there a graph map $\nu: G \rightarrow M$ so that

$$f(\nu) = \max(\alpha move_\nu) + \max(\{\beta work_p \mid p \in V(M)\} \cup \{\gamma comm_e \mid e \in E(M)\}) \leq K ?$$

A decision problem (one with a yes-no answer) is said to be in class P if there is an algorithm which solves any instance of the problem in time which can be bounded by a polynomial in the size of the instance. Such an algorithm is called "a polynomial time" algorithm.

A decision problem is said to be in class NP if for any instance for which the answer to the problem is "yes", there is a polynomial sized "certificate" which can, again in polynomial time, be used to demonstrate that the answer is "yes". Our problem is in this class because the map ν itself is the certificate.

Now, $P \subseteq NP$ since if the problem can be solved in polynomial time, we can use the problem itself as the certificate. It seems intuitively clear that NP could be much larger than P , but so far no one has been able to prove it (or to disprove it, that is, to show that $P = NP$).

³Notice that even this expression is an approximation to the exact time, since (for instance) some processors could overlap moving of vertices with computation.

⁴Originally, we called this "graph embedding", but that is used more often to refer to embedding a graph into a surface.

However, there is a nicely developed theory of “NP-completeness”. Karp, in [14], showed that there are some problems which are the hardest problems in the class NP , in the sense that if we could solve one of these problems in polynomial time, then *all* problems in NP could be solved in polynomial time. Such problems are called NP-complete; if any one of them is in P , then $P = NP$. As the latter is considered unlikely, this is good evidence that there is indeed no algorithm to solve any NP-complete problem in polynomial time.

Furthermore, to show that a decision problem is NP-complete, it suffices to show that if a polynomial time algorithm was found for that decision problem, it could be used to develop a polynomial time algorithm for some specific NP-complete problem. See [6] for details.

For our problem, this is not hard. We will reduce our problem to the following:

MINIMUM CUT INTO EQUAL PARTS

INSTANCE: Graph G , positive integer K .

QUESTION: Is there a partition of $V(G)$ into parts V_1 and V_2 such that V_1 and V_2 are of equal size, and no more than K edges pass between them?

which can be easily seen to be NP-complete by reduction to [6]’s ND17.

We can solve this problem by remapping by letting $\alpha = 0$ (so we don’t need an initial map μ), letting all weights be 1, and taking the host graph M to be the graph with two vertices connected by an edge. We will use the same value of K in both problems. Thus, the graph map ν is just a partition of $V(G)$ into two parts, and $f(\nu) = \max(\{\beta work_p \mid p \in V(M)\} \cup \{\gamma comm_e \mid e \in E(M)\})$ where now $work_p$ is just the number of vertices in each part of the partition, and $comm_e$ is the number of edges crossing between the parts.

So if we let $\beta = 2K/\#V(G)$ where $\#V(G)$ is the number of vertices of G , then only partitions into equal parts will satisfy $f(\nu) \leq K$. Setting $\gamma = 1$ completes the proof: the remapping will be possible only if at most K edges cross the partition.

2 Previous Work and Algorithms

2.1 Work on Graph Partition

The *graph partition* problem can be viewed as a specification of the graph mapping problem, in which the host graph is assumed to be complete (each vertex is connected to every other vertex). Typically, the vertex set of the graph is divided onto two equal parts V_1 and V_2 ; if more parts are desired, recursion is employed.

This problem been studied in, for example, [1,12,17,20,23], and several good heuristic algorithms have been found. We survey a few of the most popular.

2.1.1 Kernigan-Lin

The Kernigan-Lin algorithm [17] is a local improvement algorithm: given a partition, it attempts to make it better. There is a data structure [5] which allows Kernigan-Lin to run in linear time.

```

kl(  $V_1, V_2$  ) {
    best := ( $V_1, V_2$ );
    best_cost := cost(  $V_1, V_2$  );
    mark all vertices as unfrozen;
    while( there are unfrozen vertices in both  $V_1$  and  $V_2$  ) {
        transfer(  $V_1, V_2$  );
        transfer(  $V_2, V_1$  );
        if( cost( $V_1, V_2$ ) < best_cost ) {
            best := ( $V_1, V_2$ );
            best_cost := cost( $V_1, V_2$ );
        }
    }
    return best;
}

transfer(  $X, Y$  ) {
    Choose  $v \in X$  minimizing cost(  $V_1 \setminus v, V_2 \cup v$  );
     $X := X \setminus v$ ;
     $Y := Y \cup v$ ;
    mark  $v$  as frozen in  $Y$ ;
}

```

Figure 9: One step of the Kernigan-Lin graph partition improvement algorithm.

The Kernigan-Lin algorithm works by improving a partition of a graph into two parts, say V_1 and V_2 . (So $V(G) = V_1 \cup V_2$, and $V_1 \cap V_2 = \emptyset$.) The *cost* of such a partition is the number of edges between V_1 and V_2 .

One step of the algorithm begins by saving the current partition as the best-so-far, along with its cost. We then loop over the following: a vertex is transferred from V_1 to V_2 ; then a vertex is transferred back from V_2 to V_1 . The resulting partition is compared with the best-so-far, with the new partition being saved as best-so-far if it is strictly better than the old one.

The vertex to be transferred from V_1 to V_2 is the vertex v which minimizes the cost of the partition ($V_1 \setminus v, V_2 \cup v$) (regardless of whether this partition is an improvement over the previous one). That is, we move the vertex giving the best (or least bad) partition if it is transferred to the other part. (The improvement (positive or negative) is called the “gain” associated with the vertex and partition.) However, once a vertex has been transferred once, it is “frozen” and can not be transferred again. The vertex transferred the other way is chosen similarly.

This step is shown in pseudocode in figure 9; the step would be called repeatedly until the partition does not change. Note that although partitions better than the current partition are considered during the step, the returned partition is at least as good as the input partition.

An interesting variant of Kernigan-Lin is the Mob heuristic of [23]. Essentially, instead of transferring vertices one-by-one, the Mob heuristic transfers vertices in groups; the size of the groups are chosen beforehand according to a “Mob schedule”. Since it would not be cost effective to find the set of k vertices which minimize the cost of the new partition, the “mob” to be transferred is chosen to be the k vertices which would

give the k best partitions if transferred individually.

Although Mob uses the same method of deciding which vertices are good candidates for moving, there is an important difference: newly arrived vertices are not frozen, that is, vertices can move back and forth several times.

Because of the group transfer, this algorithm may be more efficient when done in parallel; however, because of the group transfer, one step may not produce as good a partition as one step of the Kernigan-Lin algorithm.

In both cases, a data structure essentially due to [5] can be used to find the best vertex (vertices) to transfer. As pointed out in [13], this data structure is a bucket sort of the nodes of the local part of the graph. By recomputing the gain of a vertex only when a neighbor is moved, this can be kept up-to-date at low cost.

2.1.2 Spectral algorithms

The problem of finding a minimum cost partition can be expressed as an integer quadratic program. Let us assume for simplicity that the number of vertices of G is even and all node and edge weights are 1; thus, we need V_1 and V_2 to have the same size.

For each vertex i , let x_i be a ± 1 variable which is 1 if $i \in V_1$ and -1 if $i \in V_2$. Then the condition that there be an equal number of vertices in each of the two parts of the partition is expressed by the linear equation $\sum_i x_i = 0$.

We write $i \sim j$ to mean vertices i and j are adjacent; then the number of edges crossing the partition is given by $\frac{1}{8} \sum_{i \sim j} (x_i - x_j)^2$, since if i and j are in the same part, $x_i = x_j$, and so $(x_i - x_j)^2 = 0$, while if i and j are in different parts, $x_i \neq x_j$, so $x_i - x_j = \pm 2$ and $(x_i - x_j)^2 = 4$. Since each edge is counted twice (once each way), this accounts for the $1/8$. Then we have

$$\sum_{i \sim j} (x_i - x_j)^2 = \sum_{i \sim j} (x_i + x_j)^2 - 2 \sum_{i \sim j} x_i x_j = \mathbf{x}^T L \mathbf{x}$$

where L is the *Laplacian Matrix* of the graph, defined by $L = [l_{ij} \mid i, j \in V]$, where l_{ii} = the degree of vertex i and for $i \neq j$, $l_{ij} = 0$ if no edge connects i and j , and $l_{ij} = -1$ if an edge connects i and j .⁵

Thus, we have the integer quadratic program

$$\begin{aligned} & \text{Minimize } \mathbf{x}^T L \mathbf{x} \\ & \text{Subject to } \mathbf{1}^T \mathbf{x} = 0 \\ & \quad x_i = \pm 1 \end{aligned} \tag{1}$$

where $\mathbf{1}$ is the vector of all 1s. However, integer programming is also NP-complete, so this does not help very much.

⁵If A is the ordinary adjacency matrix of the graph, and D the diagonal matrix of degrees, then $L = D - A$.

But, suppose we now replace the constraint that $x_i = \pm 1$ with the weaker linear constraint $\mathbf{x}^T \mathbf{x} = n$. We are left with a quadratic program, which is easy to solve:

$$\begin{aligned} & \text{Minimize } \mathbf{x}^T L \mathbf{x} \\ & \text{Subject to } \mathbf{1}^T \mathbf{x} = 0 \\ & \qquad \qquad \mathbf{x}^T \mathbf{x} = n \end{aligned} \tag{2}$$

This program can be solved particularly easily by finding the second smallest eigenvalue of L , as follows: Let the eigenvalues of L be $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ (this is called the spectrum of L). It can be shown that $\lambda_1 = 0$, and if the graph is connected, then $\lambda_1 < \lambda_2$. Furthermore, it is easy to see that if \mathbf{x} is an eigenvector with eigenvalue λ_2 (weighted such that $\mathbf{x}^T \mathbf{x} = 1$), then it solves the quadratic program (2). (See, for instance, [12] for details of all of this.)

The spectral method finds this second largest eigenvalue and a corresponding eigenvector, and then uses this solution to (2) to find an approximate solution to the integer program (1). For instance, it might take the vertices corresponding to the $n/2$ smallest components of the eigenvector and put them in V_2 , while putting the remaining (largest) $n/2$ vertices in V_1 .

For history of the spectral method, see [11]. It can be extended to partition in four or eight parts [12, 24] by using additional eigenvalues, but not beyond that; however, the spectral algorithm can be repeated on the resulting parts of the graph.

2.1.3 Multilevel Algorithms

The multilevel partitioning algorithm works by first “coarsening” the graph by merging some sets of nodes into new “supernodes”, typically by shrinking edges. The new graph is then partitioned; this partition induces a partition of the original graph, which can be improved by the Kernigan-Lin algorithm. Good results are achieved by doing several rounds of coarsening until the coarsened graph is relatively small, then partitioning by some other method (for example, spectral). Several rounds of uncoarsening and Kernigan-Linning follow [13,15,16].

Currently, multilevel methods seem to be the best algorithms for partitioning graphs well in a short amount of time. (Simulated Annealing seems to partition graphs slightly better but in much more time.)

2.2 Work on Graph Embedding

There has been a certain amount of work on “embedding” a source graph into a target graph. In general, though, there are several differences between this work and our thinking. In an embedding, the two graphs are (roughly) of equal size, and the work has concentrated on finding maps between the vertex sets of the two graphs which have small “dilation” (the distance between the images of two adjacent vertices of the source graph in the target graph; we require dilation 1), “congestion” (the maximum number of edges of the source graph mapped to a single edge of the target graph), and similar parameters.

In addition, this work tends to focus on finding algorithms to map one class of graphs (*e.g.*, grids) onto another class of graphs (*e.g.*, hypercubes) so as to always use the smallest possible member of the target class in such a way as to obtain the provably best embedding in terms of one or more of the parameters (*e.g.*, dilation no more than 2 [4]). See also [25].

This work does not appear to apply directly to our effort. First, the problems are very hard, so results are limited to special cases. Second, we do not wish to allow any dilation in our maps, but dilation of more than one is often required in an embedding. Finally, in general our guest graph is larger than the host graph, rather than the same size used in the embedding problem (though this problem could be eliminated by using a multilevel type algorithm, where the guest graph is coarsened to a graph with the same number of vertices as the host graph). The work in [9], however, extends graph embedding techniques in the direction of our problem.

It should also be noted that the term "graph embedding" is more often used to refer to drawing a graph (in some surface) so that no edges cross [8]. This work appears to have less relevance to our problem, though see section 2.3.1.

2.3 Work on Our Problem

There has not been a lot of work on our more general problem. For embedding into hypercubes, recursive bisection often works well because of the structure of the graph. [12]'s quadrisections and octasections also can be interpreted as embedding into a small cube.

2.3.1 The Flow Algorithm

A simple technique to attack the problem is for each processor in the host graph to compute its load, globally average these, and then have each processor (somehow) adjust its load to the average. One method for adjusting the load is the "flow algorithm", discussed in my report from last summer [21].

First, we run the computation for some time period with the initial embedding. Measuring the time each processor takes to run its part of the computation graph (and, potentially, how long each individual vertex of the computation graph takes) gives us a relative performance figure p_i for each processor, measured in (for example) seconds per vertex.

We can now use this information to compute how many vertices should be on each processor: each processor gets a number of vertices proportional to the number of vertices it processed in each time step (that is, the inverse of the speed). If this could be done exactly, and the speed of the processors did not change, then all processors would finish the next round of computation in exactly the same amount of time. Unfortunately, we must round to an integer number of processors, and the processors' effective speeds may change as well on a time-shared multiprocessor.

The information about how many vertices should be on each processor and the current number of vertices on that processor, give us either a number of vertices which should be sent to that processor (a demand) or a number of nodes which that processor would like to get rid of (a supply) or neither. We can arrange rounding to be done so that total supply equals total demand, though this is not really necessary.

A minimum cost flow algorithm is then used to generate a flow which matches up the supplies and demands by paths in the processor graph. To realize this flow, a processor which is sending vertices pairs with one receiving vertices, and a variant of Kernigan-Lin is used to transfer the relevant number of vertices and reoptimize the partition between those two processors. (Using the Hendrickson and Leland [13] multiway partitioning would be better, but it is not obvious how to parallelize it.)

The downside of such an algorithm is the serial dependency introduced by computing the global average (and running the flow algorithm). Serial dependencies can hamper parallel efficiency, in that most processors will be idle during the time the serial computation is running. Also, as the host graph gets larger, the serial time required here would increase.

One option would be to do the network flow in parallel. There would probably be a lot of communications, so this might actually be slower than the serial case on some multiprocessors.

Another, straight-forward, way to get around the serialization problem is for each processor to average its load with its neighbors; call this *local averaging*. This, however, has some problems of its own. Either one must accept less uniformity of the load or repeat the algorithm multiple times; in the latter case, it might be faster to do the global computation.

Heirich [10] points out another problem with local averaging. Consider the case where the host graph is 1-dimensional grid, so processor p_i is connected to p_{i-1} and p_{i+1} (except at the ends). Then (again, except at the ends) the previous algorithm will arrive at a steady state where the load u_i on processor p_i satisfies $u_i = \frac{1}{2}(u_{i-1} + u_{i+1})$. It is shown that this means u_i is a second-order solution to the Laplace partial differential equation $\frac{\partial^2 u_i}{\partial i^2} = 0$. The solutions to this PDE are $u_i = mi + b$, but the solution we are after, load balance, requires also that $m = 0$. (It should be noted that in real architectures, the presence of boundaries or wrap arounds will eventually eliminate this problem.)

The previous analysis suggested [10]’s algorithm, “Diffusion”. The basic idea is replace the Laplace equation with an equation for heat flow, $\frac{\partial \mathbf{u}}{\partial t} - \nabla^2 \mathbf{u} = 0$. An iterative solution to this equation is used as the load balancing algorithm. There are two drawbacks: the algorithm is quite complicated compared with local averaging, and it appears that we are restricted to grid-like graphs as host graphs (since otherwise it is hard to make sense of ∇). This last is probably not a big restriction for the case of PIC plasma simulations, as these graphs are indeed grid-like. (And it might be possible to embed (in the sense of [8]) the host graph into a surface and interpret the differentials there.)

Note that *any* algorithm must do at least as many communication steps as the diameter of the multiprocessor graph; see section 1.3. Due to the boundaries, in many cases local averaging will achieve this bound (for instance, on the n -cube), so it is not clear that the complicated diffusion algorithm actually is better. It would be much more attractive if it could be simplified.

3 Hardware and Software Used

Currently, the test bed is developed on an “IBM compatible” PC running Linux. Linux is a unix-like operating system developed by Linus Torvalds and an ad-hoc collection of programmers on the internet [26]. With the

GNU C++ compiler, it provides a full unix development environment, and is freely redistributable under the GNU public license.

We chose to use Linux rather than *e.g.* Windows 3.1 because, even with the addition of additional software (such as Netscape, GNU Emacs, and an X server), Windows did not provide all the features we needed, nor was it particularly stable. (In particular, Netscape and Emacs could not both run at once with my setup.) In addition, since most multiprocessors (including the SP2 we use) run some variant of unix (often with the GNU C++ compiler), Linux is more similar to the target machines.

Currently, the target multiprocessor is the IBM SP-2 at the Maui High Performance Computing Center. The SP-2 is a good machine to try these algorithms out because, while its architecture is not completely representable by a graph (because of the sophisticated switching network), it can simulate an arbitrary multiprocessor connection network (again, because of the switching network). We intend to continue the development of the test bed on the PC running Linux, then transferring the code to the SP-2; this will limit the load on the SP-2 (and our network connection).

For the network software, we are using PVM (portable virtual machine) [7], which is a publicly available message passing library. As the PVM class are hidden in C++ classes, it would be not too painful to convert to MPI, the proposed standard message passing interface if that becomes necessary. At the time the project was started, MPI was rather immature, and as MPI becomes more widespread, it is expected that PVM will run on top of MPI, so it will be almost as efficient.

We also use LEDA (Library of Efficient Discrete Algorithms) [19]; this library contains a C++ graph class which we use to implement a distributed graph structure. Each task has a LEDA graph which contains all the vertices supposed to be local to that task (and all edges between them) along with “ghosts” of all neighboring vertices. (These vertices are marked with the task where they are local).

The code is written in a literate programming style [3,22] using a literate programming tool called *Pig* of our own design. Some typical pig code is shown in the appendix. (*Pig* is fairly small: it adds a few environments to \LaTeX [18], and lets \LaTeX take care of the typesetting. This means the appearance of the code can be altered by, for instance, using a different \LaTeX document class.)

4 Design Outline

4.1 Lowest level

At the lowest level, the **Task** and **NetBuffer** classes, the test bed library allows messages to be sent to tasks using the familiar C++ << notation. For instance, instead of the following (from [7], pages 89 and 90, slightly simplified here):

```
int task_ids[ ... ], master ;
...
/* Send: */
pvm_initsend(PvmDataDefault) ;
```

```

pvm_pkint( &left, 1, 1) ;
pvm_pkint( &right, 1, 1) ;
pvm_pkdouble( &deltax2, 1, 1) ;
pvm_send( task_ids[i], 4) ;
/* Receive: */
pvm_recv( master, 4) ;
pvm_upkint( &left, 1, 1) ;
pvm_upkint( &right, 1, 1) ;
pvm_upkdouble( &deltax2, 1, 1) ;
do_something( left, right, deltax2) ;
our code would use something like

```

```

Task from(...), to(...) ;
NetBuffer n(...) ;
...
// Send:
n<<left<<right<<deltax2 ;
send( to, n, 4) ;
// Receive:
recv( from, n, 4) ;
n>>left>>right>>deltax2 ;
do_something( left, right, deltax2) ;

```

Here, the 4 is a “message id”; messages are sent and received by the id, which is an integer. In PVM, the task id is also an integer, while in the test bed, this is hidden in the **Task** class. Using a standard C++ trick, all implementations of **Task** will be able to share the same header files, so it will be possible to change the underlying message passing library by simply relinking the code.

A class can be passed in this way if, just as with **iostreams**, **operator >>** and **operator <<** are defined for the class. We don’t actually use this level directly; the following levels are built on it.

4.2 Message classes

The next level is to bundle the message id and the data into an object. The requirements of such an object are given by the **Message** virtual base class; this is a class which cannot be actually be created, but which other classes inherit from. Such classes must provide the **Message** interface. The example above could be rewritten as:

```

class AMessage : public Message {
    ...
    // The 4 will be included in a constructor here

```

```

        void DoSomething() ;
        int left, right ;
        double deltax2 ;
    } am ;
    Task to, from ;
    ...
    // Send:
    to<<am ;
    // Receive:
    from>>am ;
    am.DoSomething() ;

```

4.3 Active messages

An active message executes some code on arrival. In the current implementation, it is necessary for the receiving program to call a special routine periodically to receive the messages, and all active messages must be declared in both sender and receiver.

```

class AnActiveMessage : public ActiveMessage {
    ...
    // We need to define some routines such as:
    static ActiveMessage *Creator( NetBuffer &) ;
    // and DoSomething is now declared as:
    virtual void Arrive( const Task &from) ;
    int left, right ;
    double deltax2 ;
} aam ;
Task to ;
// Send:
to<<aam ;
// Receive this and other active messages:
ActiveMessage::ReceiveWait( net)

```

This is the level we expect to write most of our algorithms. For instance, the first cut Kernigan-Lin algorithm is started by one task send a `KLMessage` to another; this `KLMessage` is an active message which, in its `Arrive` method, sends a reply to the first task, and so on.

5 Results and Future Plans

So far, the basic part of the test bed has been written (this was mostly done over the summer on a SFRP visit to Phillips Laboratory). We are continuing to work on the algorithms; the “gains” data structure is almost working, and when it is done, Kernigan-Lin and Mob (along with some related algorithms) will be easy to write.

We intend to use the test bed to test the algorithms mentioned above (include my own flow algorithm) on a variety of graphs; we hope this will include some graphs derived from PIC plasma simulations. In addition to simulated load variations, eventually we hope to add a simple PIC simulation to give more realistic load variations. We are particularly interested in whether global algorithms like the flow algorithm actually out-perform local algorithms like local averaging and diffusion; if so, then there are tradeoffs to consider.

In the theoretical direction, recently there has been some work done on showing when problems like ours have “good” approximate algorithms. This work is related to the theory of NP-completeness. We would like to investigate whether these methods could be used to either find a (theoretically) good approximation algorithm for the remapping problem, or alternatively to show that there are no such algorithms. Just as showing the remapping problem was NP-complete was helpful (in that it suggested that a polynomial-time exact algorithm was impossible), this might help in the search for good algorithms for remapping.

6 Bibliography

- [1] S. T. Barnard and H. D. Simon, *Fast Multispectral Implementation of Recursive Spectral Bisection for Partitioning Unstructured Problems*, *Concurrency* **6** (1994), 101–117.
- [2] C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation*, Adam Hilger, New York, 1991.
- [3] P. Briggs, *Nuweb Version 0.87b: A Simple Literate Programming Tool*, distributed with *nuweb*, ND.
- [4] M. Y. Chan and F. Y. L. Chin, *On Embedding Rectangular Grids in Hypercubes*, *IEEE Transactions on Computers* **37** (1988), 1285–1288.
- [5] C. M. Fiduccia and R. M. Mattheyses, *A Linear-Time Heuristic for Improving Network Partitions*, in 19th IEEE Design Automation Conference, 1982, 175–181.
- [6] M. R. Garey and D. S. Johnson, *Computers and Intractability*, Freeman, New York, 1979.
- [7] A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manchek and V. Sunderam, *PVM: Parallel Virtual Machine, A Users' Guide and Tutorial for Networked Parallel Computing*, The MIT Press, Cambridge, MA, 1994.
- [8] J. L. Gross and T. W. Tucker, *Topological Graph Theory*, Wiley Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York, 1987.
- [9] A. K. Gupta and S. E. Hambrusch, *Load Balanced Tree Embeddings*, *Parallel Computing* **18** (1992), 595–614.
- [10] A. Heirich, *Scalable Load Balancing by Diffusion*, Caltech, preprint, 1994.
- [11] B. Hendrickson and R. Leland, *An Improved Spectral Graph Partitioning Algorithm for Mapping Parallel Computations*, Sandia National Laboratories, SAND92-1460, Albuquerque, NM, 1992.

- [12] ———, *Multidimensional Spectral Load Balancing*, Sandia National Laboratories, SAND93-0074, Albuquerque, NM, 1993.
- [13] ———, *A Multilevel Algorithm for Partitioning Graphs*, Sandia National Laboratories, SAND93-1301, Albuquerque, NM, 1993.
- [14] R. M. Karp, *Reducibility Among Combinatorial Problems*, in *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher, eds., Plenum Press, New York, 1972, 85–103.
- [15] G. Karypis and V. Kumar, *A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs*, University of Minnesota, Department of Computer Science, 95-035, Minneapolis, 1995.
- [16] ———, *Parallel Multilevel Graph Partitioning*, University of Minnesota, Department of Computer Science, 95-036, Minneapolis, 1995.
- [17] B. W. Kernigan and S. Lin, *An Efficient Heuristic Procedure for Partitioning Graphs*, Bell Systems Technical Journal **49** (1970), 291–307.
- [18] L. Lamport, *LaTeX: A Document Preparation System (Second Edition)*, Addison-Wesley.
- [19] S. Näher, *LEDA User Manual Version 3.0*.
- [20] A. Pothen, H. D. Simon and K-P. Liou, *Partitioning Sparse Matrices with Eigenvectors of Graphs*, SIAM Journal of Matrix Analysis and Applications **11** (1990), 430–452.
- [21] M. Purtil, *Static and Dynamic Graph Embedding for Parallel Programming*, Phillips Laboratory, Final Report for AFOSR Summer Faculty Research Program, 1994.
- [22] N. Ramsey, *Literate-Programming Can Be Simple and Extensible*, Department of Computer Science, Princeton University, distributed with `noweb`, 1993.
- [23] J. E. Savage and M. G. Wloka, *Parallelism in Graph-Partitioning*, Journal of Parallel and Distributed Computing **13** (1991), 257–272.
- [24] P. R. Suaris and G. Kedem, *An Algorithm for Quadrissection and Its Application to Standard Cell Placement*, IEEE Transactions on Circuits and Systems **35** (1988), 294–303.
- [25] A. S. Wagner, *Embedding All Binary Trees in the Hypercube*, Journal of Parallel and Distributed Computing **18** (1993), 33–43.
- [26] M. Welsh and L. Kaufman, *Running Linux*, O'Reilly & Associates, Sebastopol, CA, 1995.

A Network Mobile Classes

A.1 Introduction

Using the previous two chapters, we now can build classes which can be sent over the network. We will set up things as follows: (*task* is an object of class *Task*, *network* is of class *Network*, and *object* is of class *Class*, a derived class of *NetClass*)

object>>*task* send *object* to *task*;

object<<*task* receive *object* from *task* (only), overwriting old contents;

NetClass object(*task*) make a new *object* reading information from *task*;

object>>*network* send *object* to all tasks in the whole network;

object<<*network* receive *object* from any *task* in the whole network, overwriting old contents;

NetClass object(*network*) make a new *object* reading information from any task in the whole network.

A.2 File Organization

There are header and code files, each called *netclass*.

```
"netclass.h" :=
    #ifndef MRP_NETCLASS_H
    #define MRP_NETCLASS_H
    <<netclass.h header files>>
    <<NetClass declarations>>
    #endif
<<netclass.h header files>> +=
    #include "util.h"
    #include "netbuf.h"
"netclass.cc" :=
    #include "netclass.h"
    <<NetClass definitions>>
```

A.3 Class Organization

We have basically one class, *NetClass*, and it is virtual.

```
<<NetClass declarations>> +=
    class NetClass {
        protected:
            <<NetClass member variables>>
            <<NetClass pure virtual functions>>
            <<NetClass ctors,dtors,etc.>>
            <<NetClass member functions>>
    } ;
    #include "task.h"
```

⁵RCS Header: /u/purtill/riker/RCS/netclass.pig,v 0.4 1995/08/14 20:46:50 purtill Exp

A.4 Interface to Derived Classes

A.4.1 Class Tag

Here we have (unfortunately) a bit hack: each `NetClass`-derived class must have a *unique* integer associated with it to be the message tag for sending such an object. (The *class* could have more than one tag, but it needs at least one, and each *object* needs exactly one.) The tag should be between 0 and 65535 (hex `0xffff`); we will prepend a magic number (in the top 16 bits) for the actual tag used by PVM or whatever. (As we will see later, the class will set the tag by calling a constructor.) These tags will all be stored in a file called `nettags.h` to make sure they stay unique. Since PVM requires the tag to be *nonnegative*, we need the high-bit to be 0.

<<netclass.h header files>> +=

```
#include "nettags.h"
```

Note the concealed use of `tag_top`.

<<NetClass member variables>> +=

```
private:
```

```
    long tag ;
```

```
protected:
```

```
    static const long tag_top ;
```

```
    long Tag() const { return tag^tag_top ; }
```

```
    long SetTag( long ) ;
```

<<NetClass definitions>> +=

```
    const long NetClass::tag_top = 0 * 0x0ACE0000 ;
```

```
    long
```

```
    NetClass::SetTag( long new_tag) {
```

```
        long old_tag = tag^tag_top ;
```

```
        tag = 0x8FFFFFFF&(new_tag^tag_top) ;
```

```
        return old_tag ;
```

```
    }
```

A.4.2 Input/Output

We need for the derived class to be able to read its contents from a `NetBuffer`, using the following functions:

<<NetClass pure virtual functions>> +=

```
protected:
```

```
    virtual NetBuffer &Put( NetBuffer &) const =0 ;
```

```
    virtual NetBuffer &Get( NetBuffer &) =0 ;
```

```
    virtual NetBuffer &GetPut( NetBuffer &, NetBuffer::Direction) ;
```

<<NetClass definitions>> +=

```
    NetBuffer &
```

```
    NetClass::GetPut( NetBuffer &n, NetBuffer::Direction dir) {
```

```
        if( dir == NetBuffer::GET)
```

```
            return Get( n) ;
```

```
        else
```

```
            return Put( n) ;
```

```
    }
```

Not that you may override `GetPut` as well as `Get` and `Put`, and in fact `Get` and `Put` may be defined in terms of `GetPut` in the obvious way. We didn't provide these as default functions because it would then be too easy to forget to provide either function; the program would compile but then go into an infinite loop.

Put is a `const` function; it should not alter the object being passed (or, in particular, destroy it). Get should, on the other hand, destroy the previous contents if necessary. (It should *not* assume that there is always previous contents to destroy: that is, doesn't just do `delete[] foo`, do `if(foo) delete[] foo`.)

Note that the derived classes should *not*, in general, define operator `<<(Task &)`, etc.; these function will be defined for `NetClasses` later using `Get` and `Put`.

A.4.3 Constructors

On the other hand, each derived class should define constructors taking only a `NetBuffer &` or a `SubNet &` argument. This ctor should set `tag`, do any other initialization the subsequent `Get` will need (for instance, setting pointer to `NULL` so that `Get` doesn't try to free them), and then call `_init` or the `<<` operator with the original argument. (Actually, in the case of the `NetBuffer` ctor, you don't need to set `tag` before calling `_init`.)

We provide an `_init` function which just calls the `<<` operator; it will work when called in the constructor because the virtual functions will be set to the derived class's functions at that point (where as making such a call in a `NetClass` ctor would call `NetClass`'s (undefined) `Get`).

Warning: However, if the heirarchy of classes is more than two deep (`NetClass` and classes derived directly from them), you could get into trouble. In that case, the derived function at level one should provide a constructor taking no arguments, and let the bottom level constructor deal with the `NetBuffer`. That is, only one call of `_init` should be made, and that at the lowest level.

Because of this, any class derived from `NetClass` should have a `protected` constructor just like the "real" one except for the call to `_init`. Suggestion: this constructor should take a `NetClass *` as argument (and ignore it).

As an alternative to the plain ctor (which leaves the class later to call `SetTag`), we include one that takes the tag as an argument.

We also provide a do-nothing constructor and a virtual destructor.

`<<NetClass ctors, dtors, etc.>> +=`

```
protected:
    void _init( NetBuffer & ) ;
    void _init( const Task & ) ;
    void _init( const SubNet & ) ;
private:
    NetClass() {
        tag = -1 ;
    }
public:
    NetClass( long t ) {
        tag = -1 ;
        SetTag( t ) ;
    }
public:
    virtual ~NetClass() {
        tag = -99 ;
    }
```

`<<NetClass definitions>> +=`

```
void
NetClass::_init( NetBuffer &n ) {
    *this << n ;
```

```

}
void
NetClass::_init( const Task &t) {
    *this << t ;
}
void
NetClass::_init( const SubNet &n) {
    *this << n ;
}

```

A.5 In/Out Operators

A.5.1 NetBuffers

Now, all we need are the << and >> operators. First, we have the ones for putting and getting from **NetBuffers**, which do not need the tag.

```

<<NetClass member functions>> +=
public:
    NetClass &operator << ( NetBuffer &) ;
    const NetClass &operator >> ( NetBuffer &) const ;
<<NetClass definitions>> +=
    NetClass &
    NetClass::operator << ( NetBuffer &n) {
        Get( n) ;
        return *this ;
    }
    const NetClass &
    NetClass::operator >> ( NetBuffer &n) const {
        Put( n) ;
        return *this ;
    }

```

We also have the operators the other way around:

```

<<NetClass declarations>> +=
    inline NetBuffer &
    operator << ( NetBuffer &nb, const NetClass &nc) {
        nc>>nb ;
        return nb ;
    }
    inline NetBuffer &
    operator >> ( NetBuffer &nb, NetClass &nc) {
        nc<<nb ;
        return nb ;
    }

```

A.5.2 SubNets

Now, the **SubNet** versions. What we have to do is:

- create a **NetBuffer**;

- put the object into it (on put);
- put/get the NetBuffer;
- get the object from the NetBuffer (on get); and
- destroy the NetBuffer.

<<NetClass member functions>> +=

```
public:
    NetClass &operator << ( const Task & ) ;
    const NetClass &operator >> ( const Task & ) const ;
    NetClass &operator << ( const SubNet & ) ;
    const NetClass &operator >> ( const SubNet & ) const ;
```

<<NetClass definitions>> +=

```
NetClass &
NetClass::operator << ( const Task &t ) {
    NetBuffer nb ;
    recv( t, nb, tag ) ;
    Get( nb ) ;
    return *this ;
}
const NetClass &
NetClass::operator >> ( const Task &t ) const {
    NetBuffer nb ;
    Put( nb ) ;
    send( t, nb, tag ) ;
    return *this ;
}
NetClass &
NetClass::operator << ( const SubNet &n ) {
    NetBuffer nb ;
    recv( n, nb, tag ) ;
    Get( nb ) ;
    return *this ;
}
const NetClass &
NetClass::operator >> ( const SubNet &n ) const {
    NetBuffer nb ;
    Put( nb ) ;
    send( n, nb, tag ) ;
    return *this ;
}
```

Again, we add the operators going the other way:

<<NetClass declarations>> +=

```
inline const Task &
operator << ( const Task &t, const NetClass &nc ) {
    nc>>t ;
    return t ;
}
```

```

}
inline const Task &
operator >> ( const Task &t, NetClass &nc) {
    nc<<t ;
    return t ;
}
inline const SubNet &
operator << ( const SubNet &nw, const NetClass &nc) {
    nc>>nw ;
    return nw ;
}
inline const SubNet &
operator >> ( const SubNet &nw, NetClass &nc) {
    nc<<nw ;
    return nw ;
}

```

A.5.3 Tasks

We can finally write the last part of the Task class.

```

<<Task NetClass member functions declarations>> +=
public:
    NetBuffer &Put( NetBuffer &nb) const {
        return ((Task *)this)->GetPut( nb, NetBuffer::PUT) ;
    }
    NetBuffer &Get( NetBuffer &nb) {
        return GetPut( nb, NetBuffer::GET) ;
    }
    NetBuffer &GetPut( NetBuffer &, NetBuffer::Direction) ;
    Task( NetBuffer &) ;
<<Task NetClass definitions>> +=
    NetBuffer &
    Task::GetPut( NetBuffer &nb, NetBuffer::Direction dir) {
        nb.GetPutMagic( Task_TAG, dir) ;
        nb.GetPut( data->tid, dir) ;
        return nb ;
    }
    Task::Task( NetBuffer &nb) : NetClass( Task_TAG) {
        data = new TaskData ;
        Get( nb) ;
    }

```

A.6 Example: Simple

```

<<NetClass declarations>> +=
template <class T>
class Simple : public NetClass {
    T &itsValue ;

```

```

    <<Simple boilerplate>>
    <<Simple NetClass definitions>>
} ;

```

A.6.1 Boilerplate

```

<<Simple boilerplate>> +=
public:
    Simple( T &value, long t) : itsValue( value), NetClass( t) {}
    Simple( T &value) : itsValue( value), NetClass( TagOf( (T *)0)) {}
    ~Simple() {}
    operator T &() { return itsValue ; }

```

A.6.2 Network stuff

```

<<Simple NetClass definitions>> +=
public:
    virtual NetBuffer &Put( NetBuffer &nb) const {
        nb<<itsValue ;
        return nb ;
    }
    virtual NetBuffer &Get( NetBuffer &nb) {
        nb>>itsValue ;
        return nb ;
    }
    Simple( NetBuffer &nb, T &value, long t) : itsValue( value), NetClass( t) {
        Get( nb) ;
    }
    Simple( NetBuffer &nb, T &value)
        : itsValue( value), NetClass( TagOf( (T *)0)) {
        Get( nb) ;
    }
}

```

A.7 Automatic Object Reception

The stuff above is enough for classes that cooperate closely. However, also wish to support a more anarchic situation when objects are sent across the net without the other end expecting them. Classes that wish to do this should derive from `ActiveMessage` rather than `NetClass`; programs using such object have to call the function `ActiveMessage::Receive` every now and again. (If we had threads, we could have a thread just doing this, but we don't and anyway I'm not sure if `libpvm` is thread safe.) Objects arriving *via* `Receive` will be created on the free store, and the virtual member function `Arrive` will be called on them. These objects will not be deleted by the `ActiveMessage` stuff, though you could do a `delete this` in `Arrive`. (Another option is for `Arrive` to attach the new object into some data structure.)

Important: all `ActiveMessage`s have to be registered by calling `ActiveMessage::Register`. It is *not* enough to do this in the constructor, since the first object of that type might come across the network. The arguments to `ActiveMessage` are the tag of the class (that is, what tag should be recognized as meaning that an object of this class is incoming), and a pointer to a (usually static member, never plain member) function to create the new object off of the network, which should take a `NetBuffer &` and return a `ActiveMessage`

*. (Usually this can just have body `return T(n)`). It's okay to register multiple times, and even to change the function. To unregister, give `NULL` as the function pointer.

Since not all programs will want this, we'll put this into its own set of files: `activemessage.h` and `activemessage.cc`.

```
"activemessage.h" :=
    #ifndef MRP_ACTIVEMESSAGE_H
    #define MRP_ACTIVEMESSAGE_H
    <<activemessage.h header files>>
    <<ActiveMessage declarations>>
    #endif
<<activemessage.h header files>> +=
    #include "netclass.h"
"activemessage.cc" :=
    #include "activemessage.h"
    <<ActiveMessage definitions>>
```

A.7.1 Overview

```
<<ActiveMessage declarations>> +=
    class ActiveMessage : public NetClass {
        <<Table and lookup functions declarations>>
        <<Arrival and departure declarations>>
        <<Send and receive declarations>>
        <<ActiveMessage ctors,dtors,etc. declarations>>
    } ;
<<ActiveMessage definitions>> +=
    <<Table and lookup functions definitions>>
    <<Arrival and departure definitions>>
    <<Send and receive definitions>>
    <<ActiveMessage ctors,dtors,etc. definitions>>
```

Table and lookup functions We have a table associating the tags with the creator functions. Each class should use the `Register` function to inform the system of the creator function for its tag(s); the `RegisterOnly` function is a variant that will signal an error if the pointer associated with tag is anything other than `NULL` or the given function. This can be used to help in error handling. (You aren't allowed to pass it a `NULL` creator function pointer.) This error checking is somewhat silly in the absence of exceptions.

```
<<activemessage.h header files>> +=
    #include "table.h"
<<Table and lookup functions declarations>> +=
    public:
        typedef ActiveMessage *CreatorFcn( NetBuffer &) ;
        typedef CreatorFcn *CreatorFcnPtr ;
    protected:
        static GTable<long, CreatorFcn> *classTablePtr ;
        static CreatorFcnPtr lookup( long tag) ;
    public:
        static CreatorFcnPtr Register( long tag, CreatorFcnPtr) ;
        static CreatorFcnPtr RegisterOnly( long tag, CreatorFcnPtr) ;
```

We use a pointer rather than an object so that Register can be called from ctors of static objects, such as the OnStartup macros make. Fortunately, we don't need to delete the table.

```
<<Table and lookup functions definitions>> +=
    template class GTable<long,ActiveMessage::CreatorFcn> ;
    GTable<long, ActiveMessage::CreatorFcn> *ActiveMessage::classTablePtr =0 ;
    ActiveMessage::CreatorFcnPtr
    ActiveMessage::lookup( long t) {
        if( !classTablePtr)
            classTablePtr = new GTable<long, ActiveMessage::CreatorFcn> ;
        return classTablePtr->Lookup( t) ;
    }
    ActiveMessage::CreatorFcnPtr
    ActiveMessage::Register( long t, CreatorFcnPtr c) {
        if( !classTablePtr)
            classTablePtr = new GTable<long, ActiveMessage::CreatorFcn> ;
        if( !c)
            return classTablePtr->Delete( t) ;
        else
            return classTablePtr->Insert( t, c) ;
    }
    static ostream &
    operator << ( ostream &out, ActiveMessage::CreatorFcnPtr cfp) {
        return out<<"<CreatorFcnPtr_#><<(void *)cfp<<">" ;
    }
    ActiveMessage::CreatorFcnPtr
    ActiveMessage::RegisterOnly( long t, CreatorFcnPtr c) {
        if( !classTablePtr)
            classTablePtr = new GTable<long, ActiveMessage::CreatorFcn> ;
        if( c == NULL)
            Throw( c, NULL) ;
        CreatorFcnPtr old = classTablePtr->Lookup( t) ;
        if( old != NULL && old != c)
            Throw( old, NULL) ;
        old = classTablePtr->Insert( t, c) ;
        return old ;
    }
}
```

Arrival and departure functions Arrive is called after the object arrives, that is, after the new object has been created. PreDepart is called just before the object's data is put into the network; PostDepart is called after the object's data has been sent. The default for the departure functions is to do nothing; Arrival is pure virtual.

```
<<Arrival and departure declarations>> +=
protected:
    virtual void Arrive( const Task &from) = 0 ;
    virtual void PreDepart( const Task &to) ;
    virtual void PostDepart( const Task &to) ;
    virtual void PreDepart( const SubNet &to) ;
    virtual void PostDepart( const SubNet &to) ;
```

```

«Arrival and departure definitions» +=
    virtual void ActiveMessage::PreDepart( const SubNet &) {}
    virtual void ActiveMessage::PostDepart( const SubNet &) {}
    virtual void ActiveMessage::PreDepart( const Task &t) {
        PreDepart( SubNet( t)) ;
    }
    virtual void ActiveMessage::PostDepart( const Task &t) {
        PostDepart( SubNet( t)) ;
    }

```

Send and receive functions

Send There is a fixed tag used to send and receive **ActiveMessage** objects, **ActiveMessage_TAG**. The tag of the object will be passed first, then the object in question.

```

«All class names declared» +=
    class ActiveMessage ;
«Send and receive declarations» +=
    public:
        void Send( const Task &) ;
        void Send( const SubNet &sn) ;
«Send and receive definitions» +=
    void
    ActiveMessage::Send( const Task &to) {
        NetBuffer nb ;
        this->PreDepart( to) ;
        if( Tag() < 0)
            throw( NetworkError( "Trying to send with negative tag.")) ;
        nb<<Tag() ;
        this->Put( nb) ;
        send( to, nb, ActiveMessage_TAG) ;
        this->PostDepart( to) ;
    } ;
    void
    ActiveMessage::Send( const SubNet &sn) {
        NetBuffer nb ;
        this->PreDepart( sn) ;
        if( Tag() < 0)
            throw( NetworkError( "Trying to send with negative tag.")) ;
        nb<<Tag() ;
        this->Put( nb) ;
        send( sn, nb, ActiveMessage_TAG) ;
        this->PostDepart( sn) ;
    }

```

Receive Every so often, the receiving program needs to call `ActiveMessage::Receive` to receive incoming `ActiveMessage` objects. It is non-blocking, and returns the number of objects received.

<<Send and receive declarations>> +=

```
public:
class NotRegistered : public BaseError {
public:
    long tag ;
    int received_before ;
    NotRegistered( long t, int r)
        : BaseError(), tag( t), received_before( r) {}
    virtual ostream &output( ostream &os) const {
        return os<<"Tag_"<<tag<<"_not_registered_(after_"<<received_before
            <<"_successful_receptions.__"<<message<<endl ;
    }
};
static int verbose ;
static int Receive( const SubNet &) ;
static int ReceiveWait( const SubNet &) ;
```

<<Send and receive definitions>> +=

```
int ActiveMessage::verbose = 0 ;
int
ActiveMessage::Receive( const SubNet &from) {
    int received = 0 ;
    while( probe( from, ActiveMessage_TAG)) {
        Task from_task ;
        NetBuffer nb ; {
            recv( from, nb, ActiveMessage_TAG, &from_task) ;
        }
        <<Receive from nb>>
    }
    return received ;
}
int
ActiveMessage::ReceiveWait( const SubNet &from) {
    int received = 0 ;
    Task from_task ;
    NetBuffer nb ; {
        recv( from, nb, ActiveMessage_TAG, &from_task) ;
    }
    <<Receive from nb>>
    return received+Receive( from) ;
}
```

<<Receive from nb>> +=

```
long msg_tag = -1 ; {
    nb>>msg_tag ;
}
CreatorFcnPtr creator ; {
    creator = lookup( msg_tag) ;
```

```

}
if( creator) {
    if( verbose)
        cout<<"Receive[Wait]:_got_ActiveMessage_with_valid_tag_"<<msg_tag
            <<endl ;
    ActiveMessage *arriver = (*creator)( nb) ;
    arriver->Arrive( from_task) ;
    ++received ;
} else {
    Throw( NotRegistered( msg_tag, received), -received-1) ;
}

```

Constructors and stuff <<ActiveMessage ctors, dtors, etc. declarations>> +=

```

private:
    ActiveMessage() : NetClass( -1) {}
public:
    ActiveMessage( long t) : NetClass( t) {}
<<ActiveMessage ctors, dtors, etc. definitions>> +=
    /* none */

```

ULTRAFAST PROCESSES AND MODULATION IN IODINE LASERS

Wolfgang Rudolph

Associate Professor

Department of Physics and Astronomy

The University of New Mexico

Albuquerque, NM 87131

Final Report for:

Summer Research Extension Program

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington DC

and

Phillips Laboratory, Albuquerque NM

and

University of New Mexico

December 1995

ULTRAFAST PROCESSES AND MODULATION IN IODINE LASERS

Wolfgang Rudolph

Associate Professor

Department of Physics and Astronomy

The University of New Mexico

Albuquerque, NM 87131

Abstract

The dynamics of a photolytic iodine laser pumped by a short excitation pulse are a sensitive measure of the relaxation behavior of the anisotropic nonthermal velocity distribution which is formed in the photodissociation process. The measured pulse profiles can be fit with a comprehensive rate equation model to obtain relaxation rates for the speed and the angular components of the velocity distribution. This novel technique is used to determine relaxation rates for various buffer gases.

Gain tuning with a magnetic field is demonstrated in an excimer laser pumped atomic iodine laser. Under the influence of a static magnetic field of approximately 400 Gauss, simultaneous lasing on the 3-4 and 2-2 transition was observed. A computer model based on the Maxwell-Bloch equations was developed to show that, under the appropriate experimental conditions, the two transitions are coherent and beat together with a frequency of 13.8 GHz, the frequency separation of the two transitions. In addition, in order to provide optimal coupling of the two transitions, a nonlinear mirror consisting of a second harmonic crystal and the output coupler of the iodine resonator is discussed.

ULTRAFAST PROCESSES AND MODULATION IN IODINE LASERS

Wolfgang Rudolph

1. Introduction

If a molecule is photodissociated by a laser pulse whose photon energy is larger than the binding energy the photofragments are in a nonequilibrium velocity distribution. The excess energy is distributed among translational, rotational, and vibronic degrees of freedom and can lead to an excitation of the electronic system. The velocity distribution is anisotropic, peaked in the polarization direction of the exciting light for a parallel transition. Collisions among the photofragments, with unexcited molecules, and with buffer gas relax the angular distribution to an isotropic one and the mean speed to its thermalized value. The knowledge of the corresponding relaxation rates provides information about the collisional dynamics and the collisional cross sections [1].

When a magnetic field is applied to atomic iodine, the energy levels are split by the Zeeman effect. Under the influence of the magnetic field, the degeneracies of the energy levels are broken and instead of the six transitions that exist in atomic iodine due to hyperfine splitting, there are on the order of 100 possible transitions, depending on the strength of the applied magnetic field. However, it has been shown that one can still refer to the spectrum of atomic iodine under the influence of the magnetic field in terms of the zero-field transitions [2]. In this case, one can think of the effect of the magnetic field as changing the relative strengths of the zero-field transitions. This effect has been used to provide fast gain switching of iodine lasers [3]. In addition, extensive studies have been made of the gain tuning, polarization properties, and frequency tuning as a function of magnetic field in cw iodine lasers [4].

2. Intracavity Measurements of Relaxation Processes in Photolytic Iodine Lasers

2.1 Experimental

With some approximations the velocity distribution is separable and can be written as a product of two functions which depend only on the speed v and angular coordinate θ , respectively [5]

$$F(\vec{v}, t) = f(v, t)I(\theta, t). \quad (1)$$

The angular distribution I , as shown in Eq. 1, depends on θ only, if the excitation is through a nonpolarized laser. For the separable velocity distribution it was found that the mean velocity and the anisotropy relax exponentially with time constants τ_v and τ_b , respectively [5]. Such relaxation is typically measured using some kind of pump-probe experiment. A strong pump pulse excites the photolysis and the speed distribution

is probed by measuring the Doppler profile as a function of time. This can be accomplished by a frequency tunable laser in an absorption or gain measurement depending on whether the photofragment of interest emerges in the ground or excited state. In the latter the photodissociation can produce optical gain. We will show that a time resolved study of the radiation of a corresponding laser can provide the relaxation dynamics.

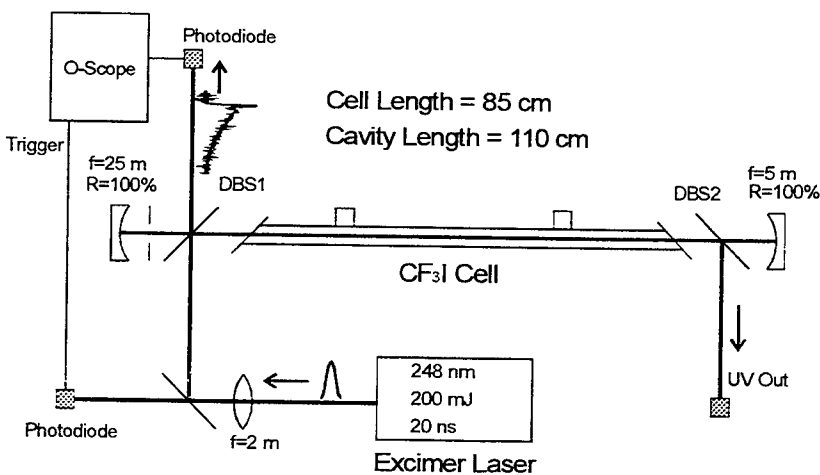


Figure 1: Photolytic laser used to determine the relaxation parameters of an anisotropic velocity distribution.

Optical gain exists in six hyper-fine transitions, the 3 - 4 transition is the strongest and is usually observed only in a laser. The experimental setup of the photolytic iodine laser used in the experiment is shown in Fig. 1.

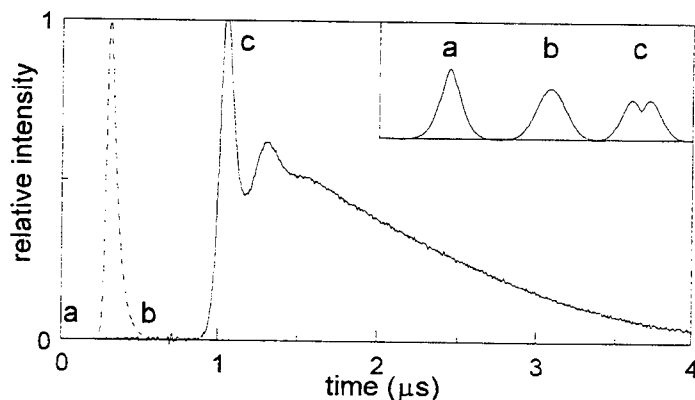
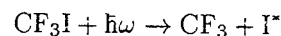


Figure 2: Typical pulse profile of a low-pressure (0.8 torr) iodine laser (solid line) pumped with 20 ns excimer pulse at $t = 0$. For comparison the results of a photolytic laser with the same passive cavity and total inversion are shown (calculated, dashed line). The inset shows a schematic of the Doppler profiles at different times in the pulse evolution.

ation by adjusting the cavity losses.

A candidate for such a photolytic reaction is the photodissociation of CF_3I



which can be used as pumping process for an iodine laser. The result of the photolysis is iodine with equal probability of being in the excited $F=2$ and $F=3$ state.

Optical gain exists in six hyper-fine transitions, the 3 - 4 transition is the strongest and is usually observed only in a laser.

A 20 ns excimer laser pulse at 248 nm pumps CF_3I in a flowing gas cell terminated by Brewster windows. The cavity consists of two 100% mirrors. Part of the iodine pulse is coupled out at a dichroic beam splitter and is displayed on a digital oscilloscope (2 ns time resolution). The pulse parameters are measured as a function of the active gas pressure and for various buffer gases. The intracavity aperture served to enforce single longitudinal and transverse mode operation by adjusting the cavity losses.

A typical output pulse is shown in Fig. 2 together with the result of a rate equation analysis that takes into account the parameters of the actual cavity and the correct total inversion density. Obviously the observed pulse occurs considerably later (pump pulse at $t = 0$) and exhibits a longer tail as what could be expected from an typical gain switched laser. The reason is the relaxation of the anisotropic and nonthermalized velocity distribution as indicated in the inset of Fig. 2. Immediately after photolysis the distribution of the angular components of the velocity is strongly anisotropic and the mean speed greater than at equilibrium which makes the gain at line center (laser frequency ω_0) larger than at equilibrium. A relatively fast relaxation to an isotropic distribution reduces the available gain at ω_0 which then slowly builds up again because of the narrowing of the Doppler profile (cooling). This relatively slow process is responsible for the delay of the onset of lasing. At the maximum of the laser pulse the gain is saturated burning a hole in the angular distribution of the velocity components (the iodine laser is linearly polarized) and causing the drop in the laser intensity. The relaxation processes filling the hole increase the gain again which manifests itself in the occurrence of a second peak and the long shoulder of the pulse.

2.2 Theoretical Model and Results

The strong dependence of the pulse features on the dynamics of the velocity relaxation suggest the possibility of fitting the data with a comprehensive rate equation model of the laser. This model is based on an effective three-level system, two upper laser levels (levels 2 and 3) linked with a cross relaxation (rate k_{23}) involving collisions with ground state iodine. Lasing occurs between level 3 and the ground state (level 1). Let us define occupation numbers of a subensemble of particles with a resonance frequency ν per frequency interval $[\nu - \frac{d\nu}{2}, \nu + \frac{d\nu}{2}]$, $\bar{N}_i(\nu - \nu_0)$, such that

$$\bar{N}_i(\nu) = \Gamma_D(\nu - \nu_0) N_i \quad (2)$$

and

$$N_i = \int_0^\infty \bar{N}_i(\nu) d\nu. \quad (3)$$

The rate equations for these occupation numbers, including relaxation between the upper hyperfine levels, and relaxation of the velocity distribution anisotropy can be written as:

$$\begin{aligned} \frac{d}{dt} \bar{N}_2 &= -q_{21} \Gamma_h(\nu - \nu_0) F \left(\bar{N}_2 - \frac{g_2}{g_1} \bar{N}_1 \right) + \frac{1.1}{\tau_b} [\bar{N}_2^{eq} - \bar{N}_2] \\ &\quad + \Gamma_D(\nu - \nu_0) k_{23} N_1 \left(N_3 - \frac{g_3}{g_2} N_2 \right) \end{aligned} \quad (4)$$

$$\frac{d}{dt} \bar{N}_1 = q_{21} \Gamma_h(\nu - \nu_0) F \left(\bar{N}_2 - \frac{g_2}{g_1} \bar{N}_1 \right) + \frac{1.1}{\tau_b} [\bar{N}_1^{eq} - \bar{N}_1] \quad (5)$$

$$\frac{d}{dt} N_3 = -\Gamma_D(\nu - \nu_0) k_{23} N_1 \left(N_3 - \frac{g_3}{g_2} N_2 \right) \quad (6)$$

where

$$\Gamma_h(\nu - \nu_0) = \frac{\Delta\nu_h}{2\pi[(\nu - \nu_0)^2 + (\Delta\nu_h/2)^2]} \quad (7)$$

is the homogeneous line shape factor, $\Gamma_D(\nu - \nu_0)$ is the (time dependent) inhomogeneous lineshape function, and

$$q_{21} = \frac{7}{12} \frac{A_{21}\lambda_0^2}{8\pi}$$

with A_{21} being the Einstein coefficient. The time τ_b describes the relaxation of the occupation numbers towards an equilibrium value $\bar{N}_i^{eq} = \Gamma_D^{eq}(\nu - \nu_0) N_i$, where $\Gamma_D^{eq}(\nu - \nu_0, t)$ is determined by the instantaneous Doppler width, which relaxes with time constant τ_v .

The corresponding rate equation for the photon flux F at $\nu = \nu_0$ reads:

$$\frac{d}{dt}F = -cq_{21}F \int_0^\infty \Gamma_h(\nu - \nu_0) \left[\bar{N}_2(\nu) - \frac{g_2}{g_1} \bar{N}_1(\nu) \right] d\nu - \frac{1.1}{\tau_r} F \quad (8)$$

where τ_r is the cavity life time.

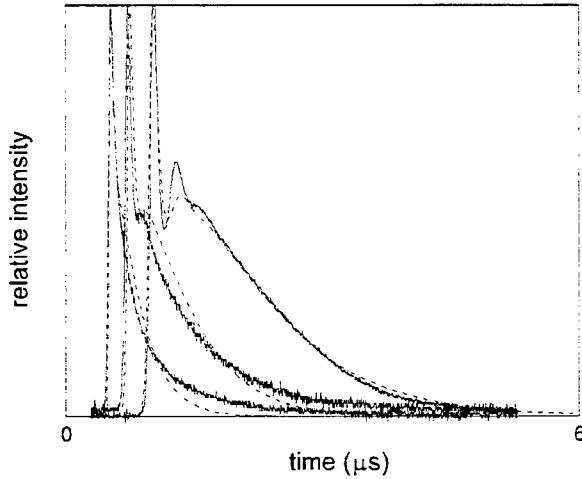


Figure 3: Iodine pulses obtained at different gas pressures and corresponding theoretical fit by a rate equation model taking into account the relaxation of an anisotropic, nonthermalized velocity distribution.

At $t = 0$ a mean velocity was calculated using the excess energy in the photolysis and distributing it among the photofragments. This speed relaxes with a time constant τ_v .

At each time step a new Γ_D^{eq} was calculated assuming a Maxwellian velocity distribution with the actual mean velocity. Any deviation of the velocity distribution from an isotropic Maxwellian distribution relaxes with a time constant τ_b .

The pulse iodine pulse profiles were measured for various pressures and buffer gases and then fitted with the above set of equations. Typical results are shown in Fig. 3. The results of such a fit are the relaxation rates for the angular component, k_b , and speed, k_v . The table below summarizes this data for several buffer gases.

Buffer Gas	k_v (10^{-11} cm ³ /molec/s)	k_b (10^{-11} cm ³ /molec/s)	k_b/k_v
Ne	1.1	3.0	2.7
Kr	5.2	11	2.1
Xe	5.2	11	2.1
N ₂	2.1	11	4.0
CO ₂	3.7	6.5	1.7

3. GHz Modulation through Magnetic Field Tuning

A sketch of the zero field gain spectrum of atomic iodine showing the relative strengths and transitions is shown in Fig. 4. The 3-4 line is the strongest transition and the one that typically lases. The second strongest is the 2-2. These two transitions are separated by a frequency of 13.8 GHz. If these two transitions are both lasing and coherent, the output intensity of the laser will be modulated with a frequency of 13.8 GHz. It is the goal of producing this modulation that is the subject of our investigations.

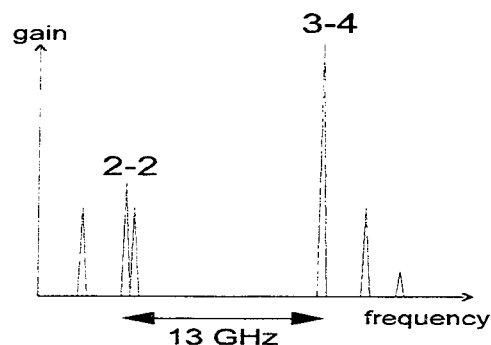


Figure 4: Spectrum of atomic iodine

In order to achieve a maximum modulation depth, the 3-4 and 2-2 transitions must be lasing with equal intensity. From Fig. 4, the 3-4 transition is more than twice as strong as the 2-2 transition with zero magnetic field; an external magnetic field is needed to equalize the gain. The effect of putting a magnetic field on the gain cell of an atomic iodine laser perpendicular to both the axis and the electric field of the laser is shown in Fig. 5. The data shown in Fig. 5 is for 20 torr of iodine and is taken from Ref. [2]. As the magnetic field increases the gain on

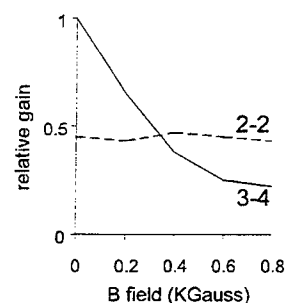


Figure 5: Gain of the 2-2 and 3-4 transitions versus magnetic field

the 3-4 line is reduced while the strength of the 2-2 line remains approximately constant. The gain of the two transitions is equalized at a magnetic field strength of approximately 400 Gauss. It is therefore in this region that we must operate our laser in order to equalize the gain of the two transitions.

3.1 Experimental Results

The experiment consisted of a photolytic iodine laser pumped by a 20 ns, 300 mJ, KrF excimer laser pulse at 248 nm. The active medium was CF_3I . Under excitation at 248 nm, CF_3I dissociates into excited state iodine (I^*) plus CF_3 . The atomic iodine then lases on the $^2P_{1/2}$ to $^2P_{3/2}$ transition with a wavelength of about $1.3 \mu\text{m}$. The laser was operated with approximately 10 torr of CF_3I . A sketch of the experimental setup is shown in Fig. 6.

Because our detection equipment is sensitive in the visible, intracavity second harmonic generation was used to convert the $1.3 \mu\text{m}$ radiation to 650 nm. With 10 torr of CF_3I in the gain cell, we obtained approximately 30 μJ of red light in a gain switched pulse several hundred ns long.

We used a fast (< 500 ps risetime) photodiode with a digital oscilloscope to view the dynamics of the fundamental and second harmonic pulses. In addition, we used a Fabry-Perot spectrometer to measure

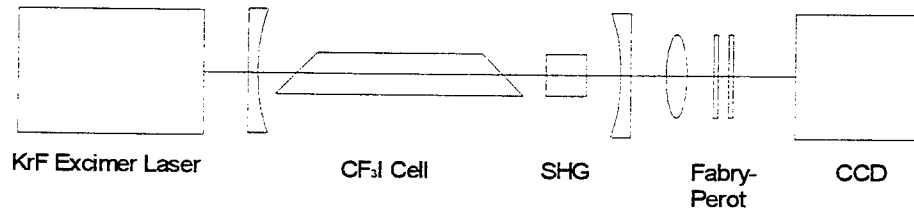


Figure 6: Sketch of the experimental setup.

the spectrum of the second harmonic. Because the pulse duration of the second harmonic was less than a microsecond, the Fabry-Perot was not scanned. Rather, we focused into the Fabry-Perot with a cylindrical lens, and then imaged the fringes that were produced transverse to the direction of propagation with a CCD camera.

A typical profile of the fringe pattern obtained from the Fabry-Perot spectrometer is shown in Fig. 7. The solid line shows the case when there is no magnetic field on the gain cell. The dotted line is the fringe pattern when a magnetic field of approximately 400 Gauss is put on the gain cell. A second set of fringes is clearly visible when there is a magnetic field present, the two sets of fringes having approximately equal intensities. From this we concluded that with a magnetic field, we obtained lasing of both the 2-2 and the 3-4 transitions.

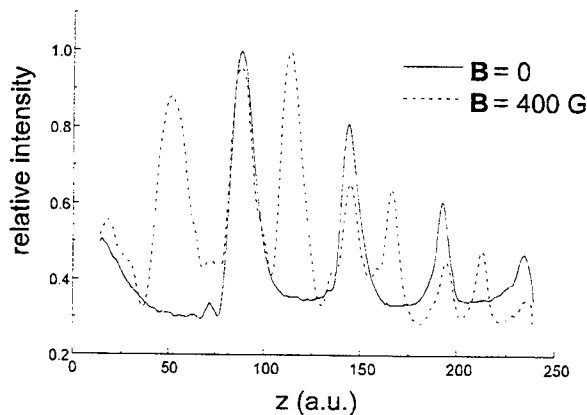


Figure 7: Fringes obtained from the Fabry-Perot spectrometer without magnetic field (solid line) and with a magnetic field of 400 Gauss (dashed line).

is propagated through the various cavity elements, such as gain medium, second harmonic crystal, outcoupler, etc. Simultaneously, a time dependent pump is introduced into the gain media. The heart of the simulation, the gain media, is based on the Maxwell-Bloch equations, as described in Ref. [6].

Preliminary results from the computer simulation indicate that a small amount of 13.8 GHz modulation could exist at high pressure (100 torr), even without a magnetic field present. More importantly, from

In order to make predictions about the expected depth of modulation under various experimental conditions, a fully coherent computer model is being developed based on the set of Maxwell-Bloch equations for the six levels of atomic iodine. In addition to modeling the development of the fast modulation in the output of the iodine laser, we plan to use the computer simulation to gain an understanding of relaxation of population and coherences among the Zeeman levels of iodine. Starting from noise (spontaneous emission), the electric field is calculated for one round trip of the laser cavity, which is assumed to be a unidirectional ring laser. Then, one cavity round trip at a time, the electric field

previous experiments with an active pressure on the order of a torr, in conjunction with the computer model, we have gained a better understanding of the importance that the kinetics of the active media play in the gain switched laser dynamics. [7]. Because the pump pulse from the excimer laser is much shorter than any processes in the gain media of the iodine laser, the pump appears as a delta function that prepares the gain in the excited state. Afterward, the population in the levels evolves through stimulated emission and collisional relaxations.

There is a relatively fast transfer of population between the upper hyperfine levels due to collisions with ground state iodine. The consequence of this is that if there is even a slight difference in the small signal gains of the two transitions, the transition that is able to build up faster will, through the collisional relaxation among the upper levels, deplete the population in the upper level of the weaker transition, decreasing its gain even further. From the computer model, a difference in gain of only 10% between the two lines can cause a difference in build up times for the two transitions of approximately 100 ns. Because we will not be able to tune the gain exactly to the crossing point shown in Fig. 5, the collisional relaxations between the upper levels will necessitate the use of an extra device to couple the energy in the two modes together.

3.2 The Nonlinear Mirror

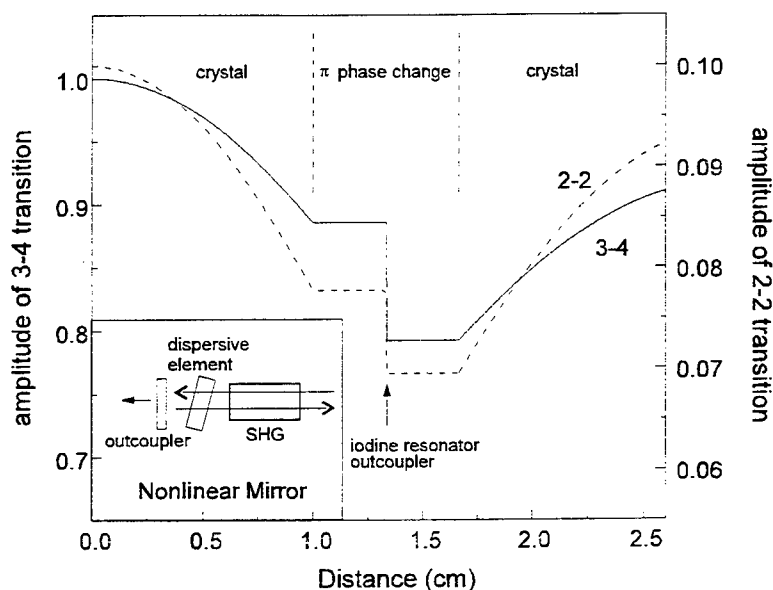


Figure 8: *The nonlinear mirror*

In order to stabilize the energy and phases of the two transitions, we have investigated the possibility of using a "nonlinear mirror" which consists of the intracavity second harmonic crystal in conjunction with a dispersive element and the cavity outcoupler. This mirror is similar to what has been used for modelocking solid state lasers [8]. Because this is a passive device, it eliminates the difficulties of driving an active

device at the required frequency of 13 GHz. A schematic of the mirror is shown in the inset of Fig. 8. The dispersive element is used to introduce a relative phase delay of π between the fundamental and sum frequency radiation. The outcoupler for the resonator has a small transmission ($\approx 10\%$) for the fundamental and 100% reflectivity for the second harmonic. In order to test the ability of the mirror to couple energy from the weaker mode into the stronger mode, we numerically solved the coupled wave equations for a single pass through the mirror. The results of this calculation, showing the field amplitudes of the 2-2 and the 3-4 transitions as a function of position are shown in Fig. 8.

As the fields pass through the mirror, they are converted into second harmonic and sum frequency fields, decreasing the amplitude of the fundamental frequencies. Then, after passing through the dispersive element and the reflecting off the outcoupler, they pass back through the second harmonic crystal. Because of the phase delay of π introduced by the dispersive element, the sum frequency radiation is reconverted back into the fundamental. At the entrance to the nonlinear mirror, the ratio of the amplitudes of the 3-4 to the 2-2 is 10; at the exit, the ratio is reduced to 9.8. This shows the ability of the nonlinear mirror to couple energy from the stronger line into the weaker line.

4. Conclusions

We developed a novel type of intracavity spectroscopy for measuring the relaxation behavior of anisotropic velocity distributions. The results for the noble gases follow to some extent the predictions from hard-sphere collision theory that heavier buffer gases tend to accelerate the relaxation processes. The data for the polyatomic gases are quite different. This is not surprising since the collisions involve also the excitation of vibrational and rotational degrees of freedom. The relaxation of the angular components of the velocity distribution is always faster than the speed relaxation. Detailed results of this project are published in [7],[9].

Through the use of a magnetic field applied to the gain cell of a photolytic iodine laser, we have obtained simultaneous lasing of two distinct hyperfine transitions. We have also developed a fully coherent computer model to help investigate the fast modulation produced in the output of the laser if the two transitions are coherent. As a result of investigations at pressures of approximately a torr of active gas and modeling with the computer simulation, we expect that collisional relaxations among the upper laser levels will act to destabilize the gains of the two transitions, even when we are operating near the crossing point where the gains are nearly equal. To this end, we have investigated the possibility of using intracavity second harmonic generation in conjunction with a dispersive element and the cavity outcoupler to form a nonlinear mirror that acts to stabilize the modes.

References

- [1] P. L. Houston. *J. Phys. Chem.*, 91:5388, 1987.
- [2] M. A. Kelly, J. K. McIver, R. F. Shea, and G. D. Hager. *IEEE J. Quantum Electron.*, QE-27:263, 1991.
- [3] V. M. Kiselev, B. D. Bodrov, A. S. Grenishin, and T. N. Kotlikova. *Sov. J. Quantum Electron.*, 9:806, 1979.
- [4] G. D. Hager, D. Kopf, B. S. Hunt, B. Anderson, C. Woolhiser, and P. Crowell. *IEEE J. Quantum Electron.*, 29:933, 1993.
- [5] J. I. Cline, C. A. Taatjes, and S. R. Leone. *J. Chem. Phys.*, 93:6543, 1990.
- [6] G. Brederlow, E. Fill, and K. J. Witte. *The high-power iodine laser*. Springer, Berlin, 1983.
- [7] J. Nicholson, W. Rudolph, J. McIver, R. Tate, and G. Hager. *Applied Physics Letters*, 67:64, 1995.
- [8] K. A. Stankov. *Appl. Phys. B*, 45:191, 1988.
- [9] G. Hager J. Nicholson, W. Rudolph. *Chemical Physics*, 1996, in print.

**COPLANAR CONICAL PLATES IN A UNIFORM DIELECTRIC LENS
WITH MATCHING CONICAL PLATES FOR FEEDING A
PARABOLOIDAL REFLECTOR**

Alexander P. Stone
Professor
Department of Mathematics and Statistics

University of New Mexico
Albuquerque, New Mexico 87131

Final Report for:
Summer Research Extension Program

Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.

December 1995

COPLANAR CONICAL PLATES IN A UNIFORM DIELECTRIC LENS WITH MATCHING CONICAL PLATES FOR FEEDING A PARABOLOIDAL REFLECTOR

Alexander P. Stone
Department of Mathematics and Statistics
University of New Mexico
Albuquerque, NM 87131

Abstract

One form of an impulse radiating antenna (IRA) consists of a paraboloidal reflector fed by conical transmission lines that propagate a spherical TEM wave, which originates from the focal point of the reflector. The design considerations of a uniform dielectric lens useful in launching a spherical TEM wave onto such a paraboloidal reflector have been investigated in earlier work and the geometrical parameters for the lens design determined. The lens design in essence consists of specifying the cylindrical radius of the outermost ray which intercepts the reflector, a dielectric medium for the lens, the angle from the apex of the conical transmission line (focal point) to the edge of the reflector, and the angle made by the line inside the lens with respect to the directions of launch.

There are several reasons for employing dielectric lenses in such an application, and these include the fact that a dielectric medium can be used to ensure a spherical TEM wave launch inside the lens. All rays emanating from the switch center should exit the lens boundary and end up on a spherical wavefront centered on the focal point at the same instant. The "equal transit-time" condition, which provides the basis for the lens design, generates an optimal spherical wave launch onto the reflector. The frequency dependence of the dielectric constant of the lens medium is also an important consideration in the choice of the material. The dielectric constant chosen is approximately 2.26, which is that of materials such as transformer oil or polyethylene and is approximately frequency independent over the frequency range of interest. In this report impedance characteristics of coplanar conical plate geometries which pass through a lens boundary are investigated. The lens impedance as a function of the half-angle of the interior conical plates, as well as the impedance outside of the lens is calculated.

COPLANAR CONICAL PLATES IN A UNIFORM DIELECTRIC LENS WITH MATCHING CONICAL PLATES FOR FEEDING A PARABOLOIDAL REFLECTOR

Alexander P. Stone

Introduction

In the design of waveguide transitions lenses we desire to transmit a TEM wave, ideally with no reflection or distortion, from one transmission line to another. Such waveguide transition regions are usually referred to as EM lenses or more specifically, transient lenses. This goal is accomplished by specifying the lens geometry and constitutive parameters (i.e., the shape and medium of the EM lens). The physical properties of these lenses, given by the permeability μ and the permittivity ϵ , may be a function of position, but we assume that these properties are frequency independent. The conductivity of the medium is taken to be zero, and cross sectional dimensions are large compared to the wavelengths at the high frequencies of interest. This is in contrast to a lens such as the Luneburg lens or Maxwell fish eye lens, both of which are based on a geometric optics approximation. The need for a low dispersion system also argues for TEM guiding structures. Since we may also wish to change the direction of propagation of a wave being transmitted from one region to another, we must also allow for distortion introduced at bends. While in many cases we can obtain exact solutions to the lens design problem, approximations are generally involved in the practical realizations of most EM lenses. For example, one may have to cut off lenses that are theoretically infinite in extent. Moreover, frequency independence may not be realized exactly. The exact solutions to design problems are usually obtained by one of two basic approaches. The first method is a differential-geometric approach, while the second method is a differential-impedance-matching and transit-time-conservation approach. In this report we are concerned with a modification of the second method as applied to a conical geometry.

We note that one form of an impulse radiating antenna (IRA) which has been studied consists of a paraboloidal reflector fed by conical transmission lines that propagate a spherical TEM wave, which originates from the focal point of the reflector. A diagram which depicts such an IRA is shown schematically in Figure 1. This type of IRA has been analyzed in the past for such performance characteristics as the prepulse or feed step and the impulse. The design considerations of a uniform dielectric lens useful in launching a spherical TEM wave onto such a paraboloidal reflector have also been investigated in earlier work[1, 2, 3], and the geometrical parameters for the lens design are indicated in Figure 2. The lens design in essence consists of specifying certain parameters. These include the cylindrical radius, h , of the outermost ray which intercepts the reflector, the relative permittivity ϵ_r of the lens medium, and the angle, $\theta_r^{(out)}$, from the apex of the conical transmission line (focal point) to the edge of the reflector. This angle is determined by the reflector geometry through the ratio F/D , where F is the focal distance from the center of the reflector and D is the

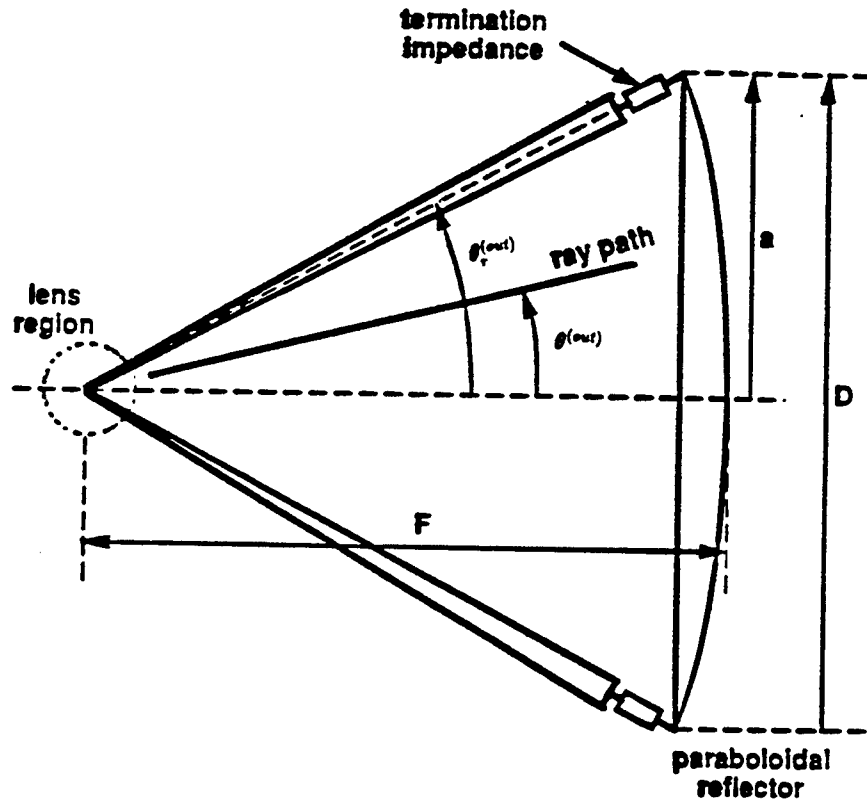


Figure 1: Schematic diagram of Prototype IRA-2

reflector diameter. We also need to specify the angle $\theta_r^{(in)}$ made by the line inside the lens with respect to the directions of launch.

There are reasons for employing dielectric lenses in such an application, and these may be briefly summarized as follows. First, the ideal point source at the focal point is physically realized by a closing switch. The high voltage between the switch electrodes dictates the use of a dielectric medium surrounding the switch enclosure. Secondly, a dielectric medium can also be used to ensure a spherical TEM wave launch inside the lens. All rays emanating from the switch center should exit the lens boundary and end up on a spherical wavefront centered on the focal point at the same instant. The "equal transit-time" condition, which provides the basis for the lens design, generates an optimal spherical wave (not truly TEM at very early times due to different transmission coefficients at different points on the lens surface) launch onto the reflector. Note that $\theta_r^{(in)}$ in [1] is taken as 90° and defines the center (symmetry plane) of the conical plate, thereby maximizing the conical-transmission line impedance inside the lens, and the plate spacing for high voltage reasons. The frequency dependence of the dielectric constant of the lens medium is also an important consideration in the choice of the material. The dielectric constant chosen is approximately 2.26, which is the dielectric constant for materials such as transformer oil or polyethylene and is approximately frequency independent over the frequency range of interest.

In this report we are concerned with impedance calculations for a lens geometry, which we take as a coplaner

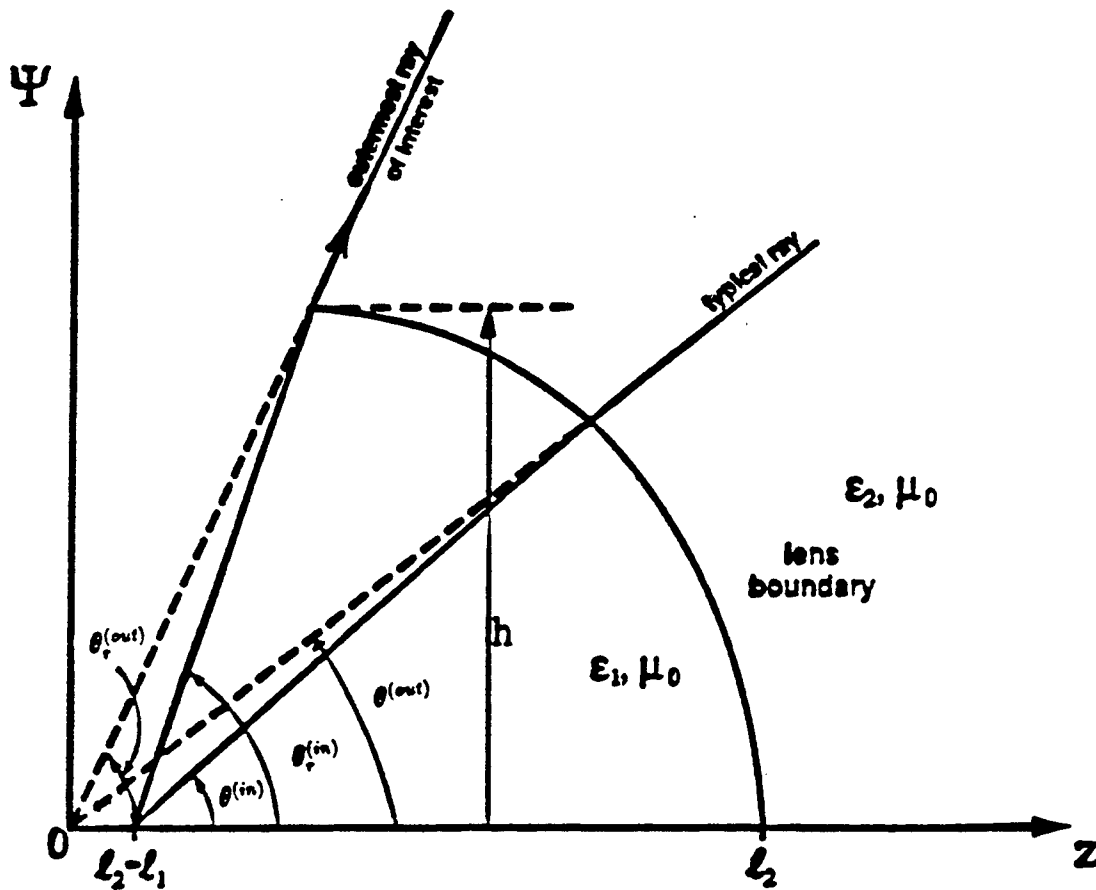


Figure 2: Lens Design Parameters

conical plate geometry as investigated in [2]. The lens impedance as a function of the half-angle, α' , (Figure 3) of the interior conical plates, as well as the impedance outside of the lens, is calculated on the basis of this coplaner conical geometry. The reflector geometry, specified by the F/D ratio, determines the acceptable range for the angle α' . The notation in Figure 3 is the same as that which appears in [2]. For consistency, changes have been made in the notation which appears in [1]. For example, θ_1 becomes $\beta_1^{(in)}$ or $\beta_2^{(in)}$, θ_2 becomes $\beta_1^{(out)}$ or $\beta_2^{(out)}$, while θ_{1max} is replaced by $\theta_r^{(in)}$ and θ_{2max} is replaced by $\theta_r^{(out)}$. The lens characteristic impedance is denoted by Z_{in} while the characteristic impedance of the region outside of the lens is denoted by Z_{out} .

Lens Design Theory

Recent work has led to the development of two types of Impulse Radiating Antenna (IRA). These two designs include a 23 cm diameter reflector IRA, with an F/D of 0.38, and a solid dielectric-lens IRA, or dielectric-immersed lens IRA. This latter design is a Transverse Electromagnetic (TEM) horn immersed in a dielectric, with a prolate spheroidal lens interface to air. The reflector IRA consists of a paraboloidal reflector, fed by

curve, which expresses the angle $\theta^{(out)}$ as a function of $\theta^{(in)}$. The resulting expression for $\cos(\theta^{(out)})$ is

$$\frac{AB \sin^2(\theta^{(in)}) + [B \cos(\theta^{(in)}) - A\epsilon_r^{\frac{1}{2}}] \sqrt{[B^2 - 2AB\epsilon_r^{\frac{1}{2}} \cos(\theta^{(in)}) + A^2\epsilon_r] - A^2 \sin^2(\theta^{(in)})}}{B^2 - 2AB\epsilon_r^{\frac{1}{2}} \cos(\theta^{(in)}) + A^2\epsilon_r} \quad (2)$$

where $A = (\ell_2/\ell_1) - 1$ and $B = (\ell_2/\ell_1) - \epsilon_r^{\frac{1}{2}}$. The parameter ℓ_2/ℓ_1 is determined once $\theta_r^{(in)}, \theta_r^{(out)}$, are ϵ_r are chosen. If $\theta_r^{(in)} = 90^\circ$, the formula for ℓ_2/ℓ_1 is, as in [1],

$$\frac{\ell_2}{\ell_1} = \epsilon_r^{\frac{1}{2}} \frac{[\cos(\theta_r^{(out)}) + \sin(\theta_r^{(out)})] - 1}{\cos(\theta_r^{(out)}) + \epsilon_r^{\frac{1}{2}} \sin(\theta_r^{(out)}) - 1} \quad (3)$$

Thus when $\theta_r^{(in)} = 90^\circ$, the parameters ℓ_2/ℓ_1 , A , and B may be expressed in terms of ϵ_r and F/D . The results are:

$$A = \frac{(\epsilon_r^{\frac{1}{2}} - 1)}{2} \left[\frac{16(F/D)^2 - 1}{4\epsilon_r^{\frac{1}{2}}(F/D) - 1} \right] \quad (4)$$

$$B = \frac{(\epsilon_r^{\frac{1}{2}} - 1)}{2} \left[\frac{16(F/D)^2 - 8\epsilon_r^{\frac{1}{2}}(F/D) + 1}{4\epsilon_r^{\frac{1}{2}}(F/D) - 1} \right] \quad (5)$$

$$\frac{\ell_2}{\ell_1} = \frac{\frac{1}{2} \left\{ (\epsilon_r^{\frac{1}{2}} - 1)[16(F/D)^2 + 8(F/D) - 1] + 2[4(F/D) - 1] \right\}}{4\sqrt{\epsilon_r}(F/D) - 1} \quad (6)$$

We may also express $\theta_r^{(out)}$ in terms of F/D by the formulas

$$\sin(\theta_r^{(out)}) = \frac{8(F/D)}{16(F/D)^2 + 1} \quad \text{and} \quad \cos(\theta_r^{(out)}) = \frac{16(F/D)^2 - 1}{16(F/D)^2 + 1} \quad (7)$$

Calculations then show that when $(F/D) > 0.337$ that $\theta^{(out)}$ attains a maximum value when $\theta^{(in)}$ exceeds 90° , since $(A\epsilon_r^{\frac{1}{2}} - B \cos(\theta^{(in)}))$ is positive. When $(F/D) < 0.337$, $\theta^{(out)}$ does not attain a maximum value for any $\theta^{(in)}$ in the range $0 \leq \theta^{(in)} \leq 180^\circ$.

The lens boundary curve is now specified by first choosing an appropriate F/D ratio which determines the angle $\theta_r^{(out)}$ of Figure 1. The lens parameters h, ϵ_r , and $\theta_r^{(in)}$ are then chosen (consistent with certain constraints which are discussed in [1]) so that the lens parameters ℓ_1/h and ℓ_2/h are obtained and a boundary curve is generated with $0 \leq \theta^{(in)}$ and $0 \leq \theta^{(out)}$. Note that $\theta_r^{(out)}$ is not the largest $\theta^{(out)}$ on the lens boundary, but that $\theta^{(out)}$ which corresponds to the edge of the reflector and similarly for $\theta_r^{(in)}$. In the cases which we investigate the angle $\theta_r^{(in)}$ will be 90° . For the chosen value of F/D , a family of lens designs may then be generated by various choices of $\theta_r^{(in)}$, subject to the constraints mentioned in [1]. In Figure 4, lens boundary curves for $\theta_r^{(in)} = 90^\circ$ and F/D values of 0.25, 0.3, 0.33, 0.4, and 0.5 are displayed. Note that larger values of F/D lead to larger lenses, while the value of $\theta_r^{(out)}$ decreases with increasing F/D . The limiting or trivial case occurs when $F/D = 0.25$. In this case, unlike the cases where $F/D > 0.25$, reflections from the lens surface (spherical) do occur, but the TEM wave is not distorted. There are also some limitations on how small $\theta^{(out)}$ can be for a given $\theta_r^{(in)}$. These are based on the slope of the lens boundary matching the ray direction.

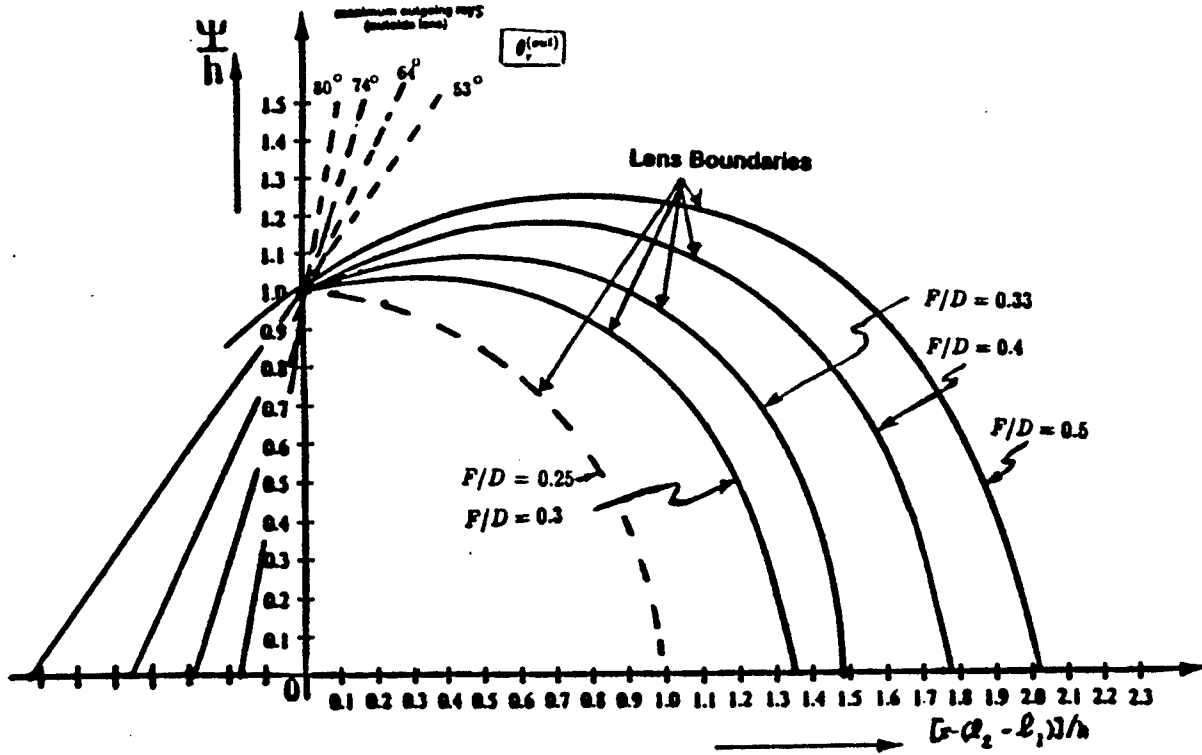


Figure 4: Lens Shapes with $\theta_r^{(in)} = 90^\circ$, $\epsilon_r = 2.26$ (Cylindrical Coordinates)

Impedance Calculations

If we consider two regions with characteristic impedances Z_1 and Z_2 and if these impedances are equal, various transfer functions and reflection coefficients may be obtained. When the impedances are unequal the bounds on the various coefficients become more complicated.

We begin by fixing the F/D ratio which determines the angle $\theta_r^{(out)}$ of Figure 1. Since the angle $\theta_r^{(in)}$ is 90° for all chosen values of F/D , and $\epsilon_r = 2.26$, the range of the angle α' is determined. Thus for all values of F/D considered, the value of Z_{in} depends only on α' . The lens characteristic impedance Z_{in} can then be calculated by the procedure described in this section. Note that $2\alpha' = \beta_2^{(in)} - \beta_1^{(in)}$ and so the value of α' determines the angles $\beta_i^{(in)}$, $i = 1, 2$. The angles $\beta_i^{(out)}$ are then found from equation (2) and thus the lens shape will be determined.

The problem of calculating the characteristic impedance of the lens region and the region exterior to the lens can now be considered. The geometry, that of coplanar conical plates, is described in Figure 3. A formula for the characteristic impedance of such a geometry appears in [2] and is given by

$$Z_{in} = \frac{Z_0}{\sqrt{\epsilon_r}} \frac{K(m_{in})}{K'(m_{in})} \quad (8)$$

where ϵ_r is the relative permittivity for the lens material and the parameter m_{in} is given by

$$m_{in} = \frac{\tan^2(\beta_1^{(in)}/2)}{\tan^2(\beta_2^{(in)}/2)} \quad (9)$$

The impedance Z_0 is 376.73 ohms. The quantities $K(m_{in})$ and $K'(m_{in})$ are complete elliptic integrals of the first kind. Formulas for these integrals appear in many places (for example, in [5]) and are given by

$$K(m_{in}) = \int_0^{\pi/2} (1 - m_{in} \sin^2(\theta))^{-\frac{1}{2}} d\theta \quad (10)$$

$$K'(m_{in}) = K(1 - m_{in}) \equiv K(m_1^{(in)}) \quad (11)$$

where $0 \leq m_{in} < 1$ and $m_1^{(in)} = 1 - m_{in}$. We may also introduce a geometric impedance factor $f_g^{(in)}$ as

$$f_g^{(in)} = \frac{K(m_{in})}{K'(m_{in})} \quad (12)$$

and thus rewrite (9) in the form

$$Z_{in} = f_g^{(in)} Z_0 / \sqrt{\epsilon_r} \quad (13)$$

Similar expressions hold for the region exterior to the lens. Thus we take for the characteristic impedance of this region

$$Z_{out} = Z_0 \frac{K(m_{out})}{K'(m_{out})} \quad (14)$$

where

$$m_{out} = \frac{\tan^2(\beta_1^{(out)}/2)}{\tan^2(\beta_2^{(out)}/2)} \quad \text{and} \quad K'(m_{out}) = K(1 - m_{out}). \quad (15)$$

Likewise a geometric factor, $f_g^{(out)}$, is given by

$$f_g^{(out)} = \frac{K(m_{out})}{K'(m_{out})}. \quad (16)$$

The elliptic integrals which depend on the parameters m_{in} and m_{out} may be evaluated by the tables in [5] or by use of the Mathematica program.

When m_{in} , or m_{out} , is near 0 or 1, there are expressions for $K(m)/K'(m)$ which may be derived. One may then verify that

$$\frac{Z_{out}}{Z_{in}} = \sqrt{\epsilon_r} \frac{K(m_{out})}{K'(m_{out})} \cdot \frac{K'(m_{in})}{K(m_{in})} \sim \sqrt{\epsilon_r} \quad (17)$$

as $m_{in} \rightarrow 1$.

Armed with the above formulas we may now calculate the impedances Z_{in} and Z_{out} , and also the ratio Z_{out}/Z_{in} for various practical values of F/D . In our impedance calculations we have chosen F/D ratios of 2.5, 3.0, 3.3, 4.0 and 5.0. Graphical results are presented in Figure 5, which gives plots of Z_{out}/Z_{in} versus α' , for each of the chosen values of F/D . The special case where $F/D = 0.25$ was mentioned in Section 2. Note that for our choice of $\epsilon_r = 2.26$ and for the limiting value of $F/D = 0.25$, we have $Z_{out}/Z_{in} = \sqrt{\epsilon_r} = 1.503$.

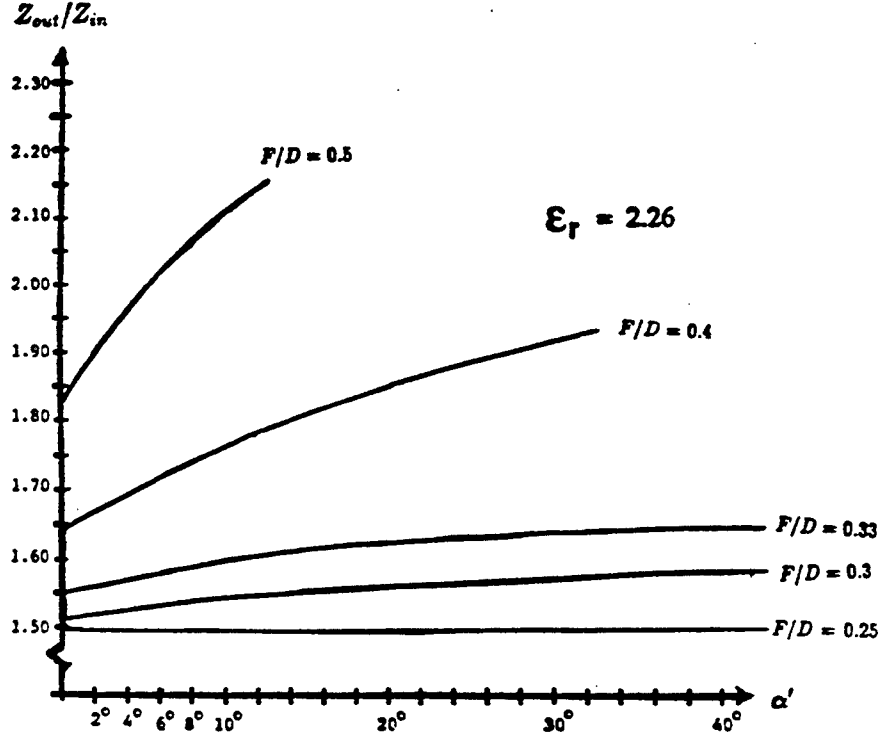


Figure 5: Impedance Ratio, Z_{out}/Z_{in} , versus half-angle, α'

Conclusions

In Figure 5 we observe that as α' increases, the ratio Z_{out}/Z_{in} increases. Small values of α' correspond to larger values of Z_{in} , and so as $\alpha' \rightarrow 0$, Z_{in} increases and smaller values of Z_{out}/Z_{in} imply a better match at the lower frequencies (radian wavelengths of the order of h or larger) for which a transmission-line approximation is appropriate. As $\alpha' \rightarrow 0$, $Z_{out}/Z_{in} \rightarrow \sqrt{\epsilon_r} = 1.503$, which is the ratio of the wave impedances of the media. For a ray propagating along the z axis (in the positive z direction) this represents a transmission coefficient of

$$T_\epsilon = 2\sqrt{\epsilon_r}/1 + \sqrt{\epsilon_r} = 1.201 \quad (18)$$

while the smaller values of α' make the transmission-line transmission coefficient

$$T_Z = \left(2 \frac{Z_0}{Z_n}\right) / \left(1 + \frac{Z_{out}}{Z_{in}}\right) \quad (19)$$

approach the same value. Note that for $F/D = 0.25$ the two transmission coefficients are the same for all α' .

The special case of $F/D = 0.25$ (a spherical lens) has the property that waves transmitted and reflected via the lens boundary are spherical TEM waves. However the reflected wave then is focused on the apex of the interior conical transmission line where it reflects back toward and through the lens boundary (with a fast rise time) unless there is a matched load at this apex. Larger F/D values make this wave reflected

Charles Swenson report unavailable at time of publication.

Improved Methods of Tilt Measurement for Extended Images in the Presence
of Atmospheric Disturbances Using Optical Flow

John I. Lipp
Doctoral Candidate
Department of Electrical Engineering

and

Jeff B. Burl
Assistant Professor
Department of Electrical Engineering

Michigan Technological University
Houghton, MI

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

Phillips Laboratory

December 1995

IMPROVED METHODS OF TILT MEASUREMENT FOR EXTENDED IMAGES IN THE PRESENCE OF ATMOSPHERIC DISTURBANCES USING OPTICAL FLOW

John I. Lipp
Doctoral Candidate
Department of Electrical Engineering
Michigan Technological University
Houghton, MI 49931

Dr. Jeff Burl
Assistant Professor
Department of Electrical Engineering
Michigan Technological University
Houghton, MI 49931

Abstract

Early tracking systems were largely based on centroid algorithms. Centroid algorithms work best when the target is *unresolved*, that is, when the features of the target are not distinguishable by the tracker's detector. Increasing the detector resolution to resolve target features results in susceptibility to variations in target intensity. In the case of atmospheric tracking, target intensity variations can be severe and tracker errors large. To mitigate tracker sensitivity to the atmosphere, Dr. John Fitts of Hughes Aircraft Company developed the *optimum weighting function correlation tracker* in the early 1970's. This algorithm, referred to as the FITTS algorithm, is not more complex than centroid algorithms, yet performs well with resolved targets.

The performance of the FITTS algorithm can be improved in at least two ways. First, the implementation of the FITTS algorithm requires the numerical estimation of image spatial gradients. Fitts used a 3-point central difference (3PCD) derivative estimator in his algorithm. It is shown to provide good performance as long as the target contains only low spatial frequencies. Derivative estimators with larger *differential bandwidths* than the 3PCD derivative estimator allow more accurate velocity estimation of targets with wider spatial frequency content. This analysis also shows that the FITTS algorithm has unstable behavior when confronted with a sinusoidal image of a spatial frequency greater than 30% of the spatial sampling rate. Second, the FITTS algorithm is a special case of *optical flow* analysis (as is the centroid, surprisingly). Viewed as an optical flow algorithm, it is possible to increase the freedom of trackable motion in the target. Specifically, the target can be allowed to have *affine* motion. The resulting *affine FITTS* algorithm allows tracking of rotation, axis skewing, and target expansion/contraction in addition to velocity. Application of the affine FITTS algorithm to simulated missile data yields some improvement in tracking accuracy.

IMPROVED METHODS OF TILT MEASUREMENT FOR EXTENDED IMAGES IN THE PRESENCE OF ATMOSPHERIC DISTURBANCES USING OPTICAL FLOW

John I. Lipp and Jeff B. Burl

1 Introduction

High temporal bandwidth, precision tracking of visible wavelength targets through long atmospheric distances poses significant problems. Precision tracking requires good optical resolving power. Good optics not only resolve the target well, they also resolve atmospheric distortions well. These distortions, called *scintillation*, occur because of the turbulent mixing of air with different temperatures. The twinkling of stars on a clear night is a typical example of the scintillation phenomenon. Classical tracking algorithms, such as centroid algorithms, require high signal-to-noise ratio images for good tracking. In particular they are highly susceptible to target intensity changes caused by scintillation. If the optical resolving power is reduced to mitigate the scintillation problem, then precision tracking is not possible. Or, if the image is temporally averaged to reduce the scintillation, precision tracking is possible, but not at high temporal bandwidths. The need for a more robust tracking algorithm was addressed by Dr. John Fitts of Hughes Aircraft Company. His algorithm, the *optimum weighting functions correlation tracker* [1][2], provides sub-pixel tracking resolution. It also is not significantly more complex than the centroid algorithms it was designed to replace. Fitts' algorithm (referred to as the FITTS algorithm in the sequel) has been shown to be successful in field demonstrations with low scintillation imagery. Success in high scintillation has been more limited.

Work in tracking algorithms has continued since the 1970's. The concept of tracking has been generalized as *optical flow*. In the most general case, optical flow concerns itself with the movement (velocity) of image intensity on a picture element (pixel) scale. Section 2 of this report shows that the FITTS algorithm is a special case of *optical flow* analysis. It is also demonstrated that the centroid is related to optical flow as well through a surprising manipulation. This suggests that further improvements in tracking under high scintillation conditions may be possible within the optical flow framework.

The optical flow problem is *ill-posed* at the pixel level. Constraints on the intensity velocity field distribution are typically imposed to *regularize* the solution. For the case of the FITTS algorithm, the velocity constraint is uniformity: every pixel of the image is assumed to move with the same velocity. The FITTS-like algorithms can be formulated that have less restrictive velocity constraints. For example, the motion can be modeled as *affine*. Affine motion allows rotation, axis skewing, and expansion/contraction of the target. The algorithm that results, the affine FITTS (A-FITTS) algorithm, is developed in section 3.

An important detail in implementing any optical flow algorithm is the numerical estimation of image derivatives. In section 4, the FITTS algorithm is analyzed with the 3-point central difference (3PCD)

derivative estimator used by Fitts. It is shown to have lackluster performance for images with spatial frequencies above 12.5% of the sampling rate. This occurs because a 3PCD derivative estimator has a very low *derivative bandwidth*. Gradient estimators with wider derivative bandwidths are analyzed and shown to yield improved performance for images with higher spatial frequencies. In a side note, it is demonstrated that the FITTS algorithm has instability with a pure sinusoidal image when the frequency is greater than 30% of the sampling rate. Simulation results with missile data are presented in section 5 followed by conclusions and recommendations in section 6.

The remainder of the introduction will present the symbol notation used in this report by summarizing the biased centroid and FITTS algorithms.

1.1 Centroid Algorithm

Denote a continuous image intensity field as $I(x, y; t)$. For practical reasons, the image field is sampled at discrete points to a $N_1 \times N_2$ grid. Often the sampling involves spatial and temporal averaging from the video camera used:

$$I[n_1, n_2; k] = \frac{1}{\Delta^2 \tau} \int_{n_1 \Delta - \Delta/2}^{n_1 \Delta + \Delta/2} \int_{n_2 \Delta - \Delta/2}^{n_2 \Delta + \Delta/2} \int_{k\tau - \tau/2}^{k\tau + \tau/2} I(x, y; t) dx dy dt \quad (1)$$

The biased centroid of the sampled images is generated from a “normal” centroid

$$\text{cen}_1[k] = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} n_1 I[n_1, n_2; k] / \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} I[n_1, n_2; k] \quad (2)$$

$$\text{cen}_2[k] = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} n_2 I[n_1, n_2; k] / \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} I[n_1, n_2; k] \quad (3)$$

by adding a DC bias term

$$\text{cen}_1^b[k] = \text{cen}_1[k] + c_1 \quad \text{cen}_2^b[k] = \text{cen}_2[k] + c_2 \quad (4)$$

The DC bias is used to force an asymmetrical object to be centered in the tracker’s field of view.

1.2 Fitts’ Algorithms

Fitts has two algorithms, both similar, which he describes as “correlation trackers” [1][2]. The second algorithm, referred to as FITTS, is [2]

$$d_{1corr}[k] = - \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \frac{\partial s[n_1, n_2; k]}{\partial n_1} (I[n_1, n_2; k] - s[n_1, n_2; k]) / \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \left[\frac{\partial s[n_1, n_2; k]}{\partial n_1} \right]^2 \quad (5)$$

$$d_{2corr}[k] = - \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \frac{\partial s[n_1, n_2; k]}{\partial n_2} (I[n_1, n_2; k] - s[n_1, n_2; k]) / \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \left[\frac{\partial s[n_1, n_2; k]}{\partial n_2} \right]^2 \quad (6)$$

has been previously demonstrated as the better of the two [3]. The algorithm outputs, $d_{1corr}[k]$ and $d_{2corr}[k]$, are the estimates of the displacement of image $I[n_1, n_2; k]$ (in pixels) from the *reference image*, $s[n_1, n_2; k]$. Generating the reference image is an important part of tracking. Fitts calls his reference image $MAP[n_1, n_2; k]$ and generates it recursively via

$$Map[n_1, n_2; i] = (1 - \rho)Map[n_1, n_2; i - 1] + \rho I[n_1, n_2; i - 1] \quad (7)$$

which is simply a low-pass filter operation on previous images. The parameter ρ determines the degree of smoothing, $\rho \in [0, 1]$. Ideally, $Map(n_1, n_2; i)$ converges to the noise-free, centered object of interest. Fitts found that using $MAP[n_1, n_2; k]$ for the reference image required corrections to the tilt calculations due to the dynamic nature of the MAP formation. The corrections are found via

$$d_{1map}[k] = d_{1map}[k - 1] + \rho d_{1corr}[k - 1] \quad (8)$$

$$d_1[k] = d_{1corr}[k] + d_{1map}[k] \quad (9)$$

where $d_{1map}[k]$ is used to compensate the calculated tilts to yield a final estimate of tilt, $d_1[k]$. Equations analogous to (8-9) apply to the n_2 -direction as well.

The partial derivatives in (5-6) must be computed numerically. Fitts used a *3-point central difference* scheme [4]

$$\frac{\partial s[n_1, n_2; k]}{\partial n_1} = \frac{s[n_1 + 1, n_2; k] - s[n_1 - 1, n_2; k]}{2} \quad (10)$$

$$\frac{\partial s[n_1, n_2; k]}{\partial n_2} = \frac{s[n_1 + 1, n_2; k] - s[n_1 - 1, n_2; k]}{2} \quad (11)$$

2 Optical Flow

The classical notion of *optical flow* is the movement of the 2-dimensional intensity when a moving 3-dimensional object is projected into 2-dimensional space. An example of this is the viewing of a 3-dimensional scene by a video camera. The 3-dimensional movement of objects in the scene produces 2-dimensional movement of the video image intensity. However, optical flow can arise from situations other than 3-D scene projections. The best definition of modern optical flow is from Luetttgen, Karl, and Willsky [5]: “the apparent velocity vector field corresponding to the observed motion of brightness patterns in successive image frames.” For a sampled image, optical flow is the movement of each pixel in the image as a function of time.

Horn and Schunk are credited with discovering the optical flow *brightness constraint* equation. It is obtained by noting that the 2-dimensional intensity at a point, $I(x, y; t)$, should not change with time as the point moves around the image. That is, the total derivative is equal to zero:

$$\frac{d I(x, y; t)}{dt} = 0 \quad (12)$$

Expanding (12) with the chain rule yields

$$\frac{\partial I(x, y; t)}{\partial x} \frac{dx}{dt} + \frac{\partial I(x, y; t)}{\partial y} \frac{dy}{dt} + \frac{\partial I(x, y; t)}{\partial t} = 0 \quad (13)$$

Now define $v_x = \frac{dx}{dt}$ and $v_y = \frac{dy}{dt}$. Then (13) can be rewritten

$$\left[\frac{\partial I(x, y; t)}{\partial x} \quad \frac{\partial I(x, y; t)}{\partial y} \right] \cdot \begin{bmatrix} v_x \\ v_y \end{bmatrix} + \frac{\partial I(x, y; t)}{\partial t} = 0 \quad (14)$$

This is the classical optical flow equation. It is a constraint on the image derivatives and the x - and y -direction optical flow velocities. Note that the optical flow equation does not completely constrain the optical flow. For every two velocity components (one each for the x - and y -directions) there is only one constraint. The optical flow equation alone makes computing the optical flow an *ill-posed* problem. *Regularization* is required to make the problem well-posed.

Numerous regularization techniques that have been applied to solve (14). One of the more common is to force the motion to be smooth. Horn and Schunk proposed a smoothness constraint by noting that minimizing the Laplacian of the optical flow gave a smooth motion field [6]. On the other extreme are complicated techniques such as that by Luetttgen, et. al. They compute the optical flow on multiple scales using a Kalman filter [5]. The Kalman filter allows the most appropriate scale to be chosen for different image regions by examining the Kalman co-variance matrix.

Next it will be shown that the FITTS algorithm is a special case of optical flow using a uniformity constraint on the motion and a particular temporal gradient.

2.1 FITTS Algorithm

The FITTS algorithm can be derived many ways. Fitts uses a matched filter approach [1], while Brunson, Grockett, and Riker employ a minimum minimum mean square error criterion [7]. A maximum-likelihood approach can be used if the noise model is additive white Gaussian because the solution is also the minimum mean square error solution [8]. A minimum mean square error approach will be shown and comparisons made to the optical flow equation and centroid equations.

Assume the displacements¹ between successive images are spatially uniform and constant in time:

$$v_x(n_1, n_2; k) = d_1 \quad v_y(n_1, n_2; k) = d_2 \quad (15)$$

Then the relationship between sampled images is

$$I[n_1, n_2; k + 1] = I[n_1 - d_1, n_2 - d_2; k] \quad (16)$$

¹The term velocities is equally descriptive since a time step is involved, and was the terminology used by Fitts. From here, displacement and velocity will be both used based on which is most appropriate to the context.

Linearize the right-hand side using a Taylor series expansion

$$I[n_1, n_2; k+1] = I[n_1, n_2; k] - \frac{\partial I[n_1, n_2; k]}{\partial n_1} d_1 - \frac{\partial I[n_1, n_2; k]}{\partial n_2} d_2 + \text{h.o.t.} \quad (17)$$

Subtract the right-hand side from the left to get

$$\Delta I[n_1, n_2; k] + \frac{\partial I[n_1, n_2; k]}{\partial n_1} d_1 + \frac{\partial I[n_1, n_2; k]}{\partial n_2} d_2 = \text{h.o.t.} \quad (18)$$

where $\Delta I[n_1, n_2] = I[n_1, n_2; k+1] - I[n_1, n_2; k]$ is a very simple temporal derivative estimator².

If the higher order terms are small, (18) is a discrete optical flow equation with a specific temporal derivative. The velocities which minimize the higher order terms are found by forming the squared sum of the higher order terms over the entire image

$$\varepsilon = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \text{h.o.t.} = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{ \Delta I[n_1, n_2] + g_1[n_1, n_2]d_1 + g_2[n_1, n_2]d_2 \}^2 \quad (19)$$

where

$$g_1[n_1, n_2] \equiv \frac{\partial I[n_1, n_2; k]}{\partial n_1}, \quad g_2[n_1, n_2] \equiv \frac{\partial I[n_1, n_2; k]}{\partial n_2}$$

Next take partial derivatives vs. d_1 and d_2 :

$$\frac{\partial \varepsilon}{\partial d_1} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{ \Delta I[n_1, n_2] + g_1[n_1, n_2]d_1 + g_2[n_1, n_2]d_2 \} g_1[n_1, n_2] \quad (20)$$

$$\frac{\partial \varepsilon}{\partial d_2} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{ \Delta I[n_1, n_2] + g_1[n_1, n_2]d_1 + g_2[n_1, n_2]d_2 \} g_2[n_1, n_2] \quad (21)$$

Setting both derivatives equal to zero and solving simultaneously for d_1 and d_2 yields the FITTS correlation tracker equations

$$d_1 = \frac{(\sum \sum g_2^2) (\sum \sum \Delta I g_1) - (\sum \sum g_1 g_2) (\sum \sum \Delta I g_2)}{(\sum \sum g_1 g_2)^2 - (\sum \sum g_1^2) (\sum \sum g_2^2)} \quad (22)$$

$$d_2 = \frac{(\sum \sum g_1^2) (\sum \sum \Delta I g_2) - (\sum \sum g_1 g_2) (\sum \sum \Delta I g_1)}{(\sum \sum g_1 g_2)^2 - (\sum \sum g_1^2) (\sum \sum g_2^2)} \quad (23)$$

The $[n_1, n_2]$ indexes of ΔI , g_1 , g_2 , and the summations have been left off for compactness. Immediately, (22) and (23) do not appear to be the FITTS equations. The difference between (22-23) and (5-6) is the inclusion of *gradient cross-terms*, $\sum \sum g_1 g_2$. Fitts noticed the gradient cross-terms tend to be small and ignored them in his algorithm and its analysis. Experimental results in [9] showed that the terms are worth

²This fact has been noticed by recent researchers in OF, and this leads to the possible idea of using a different gradient scheme within the framework of the FITTS algorithm.

including to increase accuracy. Fitts also observed that only the $I[n_1, n_2; k]$ part of ΔI contributes in the summations $\sum \sum \Delta I g_1$ and $\sum \sum \Delta I g_2$. When all the negligible terms are eliminated (22) reduces to

$$d_1 = \frac{1}{\sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} g_1^2[n_1, n_2]} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} g_1[n_1, n_2] I[n_1, n_2; k] \quad (24)$$

Rewriting the n_1 -direction centroid calculation (2) to show some similarities gives

$$\text{cen}_1[k] = \frac{1}{\sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} I[n_1, n_2; k]} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} n_1 I[n_1, n_2; k] \quad (25)$$

Fitts noticed both algorithms have the form

$$y[k] = C_1 + \frac{1}{C_2} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} w[n_1, n_2] I[n_1, m; k] \quad (26)$$

where C_1, C_2 are scaling constants and $w[n_1, n_2]$ is a weighting function. Fitts proved that his algorithm minimizes the tracker's error variance of all algorithms with the form (26). Hence his name for the algorithm, the *optimum weighting function* correlation tracker.

2.2 Optical Flow to Centroid Conversion

FITTS demonstrated that his correlation tracker equations could be reduced to a form similar to a centroid, but with superior tracking capabilities. A new relationship between optical flow and centroids has been developed and is presented to show a link between optical flow and the centroid.

To develop it, first compute the temporal partial derivative of the n_1 -centroid. This yields

$$\frac{\partial \text{cen}_1[k]}{\partial k} = \left[\overline{I[k]} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} n_1 \frac{\partial I[n_1, n_2; k]}{\partial k} - \overline{In_1[k]} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \frac{\partial I[n_1, n_2; k]}{\partial k} \right] / \overline{I[k]}^2 \quad (27)$$

where

$$\overline{I[k]} = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} I[n_1, n_2; k], \quad \overline{In_1[k]} = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} n_1 I[n_1, n_2; k] \quad (28)$$

Now Combine (27) with the optical flow equation (14) to form

$$\frac{\partial \text{cen}_1[k]}{\partial k} = \left\{ \overline{In_1[k]} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \frac{\partial I[n_1, n_2; k]}{\partial [n_1, n_2]} \cdot \begin{bmatrix} d_1[n_1, n_2; k] \\ d_2[n_1, n_2; k] \end{bmatrix} - \overline{I[k]} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} n_1 \frac{\partial I[n_1, n_2; k]}{\partial [n_1, n_2]} \cdot \begin{bmatrix} d_1[n_1, n_2; k] \\ d_2[n_1, n_2; k] \end{bmatrix} \right\} / \overline{I[k]}^2 \quad (29)$$

The result is the incremental centroid change based on the optical flow. It is worth noting that if the optical flow is uniform ($d_1[n_1, n_2; k] = d_1$ and $d_2[n_1, n_2; k] = d_2$), it is not preserved by (29). That is, $\frac{\partial \text{cen}_1[k]}{\partial k} \neq d_1$.

3 The Affine FITTS Algorithm

It has been shown that the FITTS algorithm is a special case of optical flow based on a uniform motion constraint. However, a uniform motion constraint is not the only possibility within the FITTS algorithm framework. For example, the optical flow could be *affine*. An affine optical flow model (see Appendix A) allows motions that are rotations, dilations, skews, and translations (as well as any combination thereof). To proceed with the derivation, remember the goal is to find the motion that minimizes the total square matching error

$$\varepsilon = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{I[n_1, n_2; k+1] - I[n_1 - d_1[n_1, n_2], n_2 - d_2[n_1, n_2]; k]\}^2 \quad (30)$$

To make the problem tractable, a truncated Taylor series expansion of $I[n_1 - d_1[n_1, n_2], n_2 - d_2[n_1, n_2]; k]$

$$I[n_1 - d_1[n_1, n_2], n_2 - d_2[n_1, n_2]; k] = I[n_1, n_2; k] - g_1[n_1, n_2]d_1[n_1, n_2] - g_2[n_1, n_2]d_2[n_1, n_2] \quad (31)$$

is substituted in (30). Then

$$\varepsilon = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{I[n_1, n_2] - s[n_1, n_2] + g_1[n_1, n_2]d_1[n_1, n_2] + g_2[n_1, n_2]d_2[n_1, n_2]\}^2 \quad (32)$$

Now, affine motion has the form

$$d_1[n_1, n_2] = a_{11}n_1 + a_{12}n_2 + b_1 \quad (33)$$

$$d_2[n_1, n_2] = a_{21}n_1 + a_{22}n_2 + b_2 \quad (34)$$

Substituting (33-34) for the general motion in (32) gives

$$\varepsilon = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{\Delta I[n_1, n_2] + g_1[n_1, n_2](a_{11}n_1 + a_{12}n_2 + b_1) + g_2[n_1, n_2](a_{21}n_1 + a_{22}n_2 + b_2)\}^2 \quad (35)$$

Since affine motion is parameterized by 6 coefficients there are 6 partial derivatives of interest:

$$\frac{\partial \varepsilon}{\partial a_{11}} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{\Delta I[n_1, n_2] + g_1[n_1, n_2](a_{11}n_1 + a_{12}n_2 + b_1) + g_2[n_1, n_2](a_{21}n_1 + a_{22}n_2 + b_2)\} g_1[n_1, n_2]n_1 \quad (36)$$

$$\frac{\partial \varepsilon}{\partial a_{12}} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{\Delta I[n_1, n_2] + g_1[n_1, n_2](a_{11}n_1 + a_{12}n_2 + b_1) + g_2[n_1, n_2](a_{21}n_1 + a_{22}n_2 + b_2)\} g_1[n_1, n_2]n_2 \quad (37)$$

$$\frac{\partial \varepsilon}{\partial b_1} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{\Delta I[n_1, n_2] + g_1[n_1, n_2](a_{11}n_1 + a_{12}n_2 + b_1) + g_2[n_1, n_2](a_{21}n_1 + a_{22}n_2 + b_2)\} g_1[n_1, n_2] \quad (38)$$

$$\frac{\partial \varepsilon}{\partial a_{21}} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{ \Delta I[n_1, n_2] + g_1[n_1, n_2](a_{11}n_1 + a_{12}n_2 + b_1) + g_2[n_1, n_2](a_{21}n_1 + a_{22}n_2 + b_2) \} g_2[n_1, n_2]n_1 \quad (39)$$

$$\frac{\partial \varepsilon}{\partial a_{22}} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{ \Delta I[n_1, n_2] + g_1[n_1, n_2](a_{11}n_1 + a_{12}n_2 + b_1) + g_2[n_1, n_2](a_{21}n_1 + a_{22}n_2 + b_2) \} g_2[n_1, n_2]n_2 \quad (40)$$

$$\frac{\partial \varepsilon}{\partial b_2} = 2 \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \{ \Delta I[n_1, n_2] + g_1[n_1, n_2](a_{11}n_1 + a_{12}n_2 + b_1) + g_2[n_1, n_2](a_{21}n_1 + a_{22}n_2 + b_2) \} g_2[n_1, n_2] \quad (41)$$

Solving for the affine coefficients $\{a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2\}$ is not nearly as easy as solving for $\{d_1, d_2\}$ in the FITTS algorithm. The solution can be obtained by solving an $Ax = b$ equation where

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad x = \begin{bmatrix} a_{11} \\ a_{12} \\ b_1 \\ a_{21} \\ a_{22} \\ b_2 \end{bmatrix}, \quad b = - \begin{bmatrix} \sum \sum \Delta I g_1 n_1 \\ \sum \sum \Delta I g_1 n_2 \\ \sum \sum \Delta I g_1 \\ \sum \sum \Delta I g_1 n_1 \\ \sum \sum \Delta I g_1 g_2 \\ \sum \sum \Delta I g_1 \end{bmatrix}$$

$$A_{11} = \begin{bmatrix} \sum \sum g_1^2 n_1^2 & \sum \sum g_1^2 n_1 n_2 & \sum \sum g_1^2 n_1 \\ \sum \sum g_1^2 n_1 n_2 & \sum \sum g_1^2 n_2^2 & \sum \sum g_1^2 n_2 \\ \sum \sum g_1^2 n_1 & \sum \sum g_1^2 n_2 & \sum \sum g_1^2 \end{bmatrix}$$

$$A_{12} = A_{21} = \begin{bmatrix} \sum \sum g_1 g_2 n_1^2 & \sum \sum g_1 g_2 n_1 n_2 & \sum \sum g_1 g_2 n_1 \\ \sum \sum g_1 g_2 n_1 n_2 & \sum \sum g_1 g_2 n_2^2 & \sum \sum g_1 g_2 n_2 \\ \sum \sum g_1 g_2 n_1 & \sum \sum g_1 g_2 n_2 & \sum \sum g_1 g_2 \end{bmatrix}$$

$$A_{22} = \begin{bmatrix} \sum \sum g_2^2 n_1^2 & \sum \sum g_2^2 n_1 n_2 & \sum \sum g_2^2 n_1 \\ \sum \sum g_2^2 n_1 n_2 & \sum \sum g_2^2 n_2^2 & \sum \sum g_2^2 n_2 \\ \sum \sum g_2^2 n_1 & \sum \sum g_2^2 n_2 & \sum \sum g_2^2 \end{bmatrix}$$

Note that the FITTS derivatives could also be put into an $Ax = b$ form and solved to get (22) and (23). Unfortunately, in the affine FITTS case, the required matrix inverse is for a 6×6 matrix instead of a (trivial) 2×2 inverse.

4 Gradients in the FITTS Algorithm

The FITTS algorithm requires the estimation of spatial gradients in the n_1 - and n_2 -directions, as well as the estimation of a temporal gradient. Fitts selects the spatial gradient estimator for his algorithm for no more reason than it is the most common such estimator and provided good results in the tracker [2]. Fitts' analysis of his spatial gradient choice demonstrated that performance approaches the theoretical maximum when the images are quantized with a fine sampling grid. The following greatly expands on Fitts' analysis and provides some analytical results about the FITTS algorithm performance.

4.1 Analytical SINE Image

To begin the analysis of gradients in the FITTS algorithm, a benchmark is going to be established. Consider the continuous 2-dimensional, separable sine wave image

$$\begin{aligned} I(x, y; t) &= \sin(2\pi f_x [x - V_x t]) \sin(2\pi f_y [y - V_y t]) \\ &= \frac{1}{2} \{ \cos(2\pi [f_x V_x - f_y V_y] t - 2\pi [f_x x - f_y y]) - \cos(2\pi [f_x V_x + f_y V_y] t - 2\pi [f_x x + f_y y]) \} \end{aligned} \quad (42)$$

where the first form clearly shows the spatial frequencies and the second form clearly shows the temporal frequencies. The variables f_x and f_y are the spatial frequencies and V_x and V_y are the spatial velocities in the x - and y -directions, respectively. The sampled version of this analog signal, using a camera averaging model and a 1 Hz spatial sampling frequency, is

$$\begin{aligned} I[n_1, n_2; t] &= \int_{n_1-1/2}^{n_1+1/2} \int_{n_2-1/2}^{n_2+1/2} \sin(2\pi f_x [x - V_x t]) \sin(2\pi f_y [y - V_y t]) dx dy \\ &= \frac{\sin(\pi f_x)}{\pi f_x} \frac{\sin(\pi f_y)}{\pi f_y} \sin(2\pi f_x [n_1 - V_x t]) \sin(2\pi f_y [n_2 - V_y t]) \end{aligned}$$

The analog and sampled waveforms are similar within a scaling constant. The scaling constant represents the affect of sinc filtering in the x - and y -directions from the (camera's) averaging filter. Thus, in the following analysis, the image data will be

$$I[n_1, n_2; k] = \sin(2\pi f_1[n_1 - V_1 k]) \sin(2\pi f_2[n_2 - V_2 k]) \quad (43)$$

since the affect of averaging and sampling just scales the underlying analog image. The image (43) will be referred to as SINE.

Now, consider the FITTS algorithm in the continuous domain. Replace the summations with integrals and all of the numerical gradients with partial derivatives (the most ideal tracking situation). These partial derivatives are

$$\frac{\partial I(x_1, x_2; t)}{\partial x_1} = 2\pi f_1 \cos(2\pi f_1[x_1 - V_x t]) \sin(2\pi f_2[x_2 - V_y t]) \quad (44)$$

$$\frac{\partial I(x_1, x_2; t)}{\partial x_2} = 2\pi f_2 \sin(2\pi f_1[x_1 - V_x t]) \cos(2\pi f_2[x_2 - V_y t]) \quad (45)$$

$$\begin{aligned} \frac{\partial I(x_1, x_2; t)}{\partial t} = & -2\pi f_1 V_x \cos(2\pi f_1[x_1 - V_x t]) \sin(2\pi f_2[x_2 - V_y t]) \\ & -2\pi f_2 V_y \sin(2\pi f_1[x_1 - V_x t]) \cos(2\pi f_2[x_2 - V_y t]) \end{aligned} \quad (46)$$

giving rise to the integrals

$$\int_{x_1} \int_{x_2} \left(\frac{\partial I(x_1, x_2; t)}{\partial x_1} \right)^2 dx_1 dx_2 = (2\pi f_1)^2 \int_{x_1} \cos^2(2\pi f_1[x_1 - V_x t]) dx_1 \int_{x_2} \sin^2(2\pi f_2[x_2 - V_y t]) dx_2 \quad (47)$$

$$\int_{x_1} \int_{x_2} \left(\frac{\partial I(x_1, x_2; t)}{\partial x_2} \right)^2 dx_1 dx_2 = (2\pi f_2)^2 \int_{x_1} \sin^2(2\pi f_1[x_1 - V_x t]) dx_1 \int_{x_2} \cos^2(2\pi f_2[x_2 - V_y t]) dx_2 \quad (48)$$

$$\begin{aligned} \int_{x_1} \int_{x_2} \frac{\partial I(x_1, x_2; t)}{\partial x_1} \cdot \frac{\partial I(x_1, x_2; t)}{\partial x_2} dx_1 dx_2 = & (2\pi f_1)(2\pi f_2) \int_{x_1} \cos(2\pi f_1[x_1 - V_x t]) \sin(2\pi f_1[x_1 - V_x t]) dx_1 \\ & \times \int_{x_2} \cos(2\pi f_2[x_2 - V_y t]) \sin(2\pi f_2[x_2 - V_y t]) dx_2 \end{aligned} \quad (49)$$

$$\begin{aligned} \int_{x_1} \int_{x_2} \frac{\partial I(x_1, x_2; t)}{\partial t} \frac{\partial I(x_1, x_2; t)}{\partial x_1} dx_1 dx_2 = & -(2\pi f_1)^2 V_x \int_{x_1} \cos^2(2\pi f_1[x_1 - V_x t]) dx_1 \times \\ & \int_{x_2} \sin^2(2\pi f_2[x_2 - V_y t]) dx_2 - (2\pi f_1)(2\pi f_2) V_y \times \\ & \int_{x_1} \cos(2\pi f_1[x_1 - V_x t]) \sin(2\pi f_1[x_1 - V_x t]) dx_1 \times \\ & \int_{x_2} \cos(2\pi f_2[x_2 - V_y t]) \sin(2\pi f_2[x_2 - V_y t]) dx_2 \end{aligned} \quad (50)$$

$$\begin{aligned}
\int_{x_1} \int_{x_2} \frac{\partial I(x_1, x_2; t)}{\partial t} \frac{\partial I(x_1, x_2; t)}{\partial x_2} dx_1 dx_2 = & -(2\pi f_2)^2 V_y \int_{x_1} \sin^2(2\pi f_1[x_1 - V_x t]) dx_1 \times \\
& \int_{x_2} \cos^2(2\pi f_2[x_2 - V_y t]) dx_2 - (2\pi f_1)(2\pi f_2) V_x \times \\
& \int_{x_1} \cos(2\pi f_1[x_1 - V_x t]) \sin(2\pi f_1[x_1 - V_x t]) dx_1 \times \\
& \int_{x_2} \cos(2\pi f_2[x_2 - V_y t]) \sin(2\pi f_2[x_2 - V_y t]) dx_2
\end{aligned} \tag{51}$$

Substitution of these evaluated integrals into the continuous FITTS equations [2] results in $d_1 = V_x$ and $d_2 = V_y$. Therefore the FITTS algorithm is capable of finding the exact motion for the SINE image regardless of spatial frequency or image velocity when provided with the true gradients. Going to a sampled image (summations instead of integrals) and perfectly sampled derivatives does not change this result. Thus, any degradation in the FITTS algorithm's performance for the SINE image can be traced to the imperfect numerical estimation of the image's gradients. In the subsequent sections it will be shown that performance degradation can be substantial for high spatial frequencies.

4.2 Numerical Derivative Estimation

The abstract concept of a derivative and its frequency response are very simple. A first-order, one-dimensional derivative is a linear operator which has the effect of scaling a signal's Fourier transform by $j\omega$. With the simple frequency domain representation $H(\omega) = j\omega$, it is tempting to derive and analyze numerical derivative estimators in the frequency domain. The common derivative estimators, however, are derived with a different approach which will follow.

4.2.1 3-Point Central Difference

The most common gradient estimator is the non-causal, digital filter

$$h(z) = \frac{1}{2}z - \frac{1}{2}z^{-1} \tag{52}$$

referred to as a 3-point central difference (3PCD) filter [4]. Its frequency and phase responses are easily derived and plotted. Substituting $z = e^{j\omega}$ gives

$$H(\omega) = \frac{1}{2}e^{j\omega} - \frac{1}{2}e^{-j\omega} = j \sin(\omega)$$

Immediately noticeable is that the phase of this gradient estimator is exactly that desired. This leaves the magnitude to be examined. It is shown in Figure 1.a. Clearly, (52) is a good gradient estimator only at low frequencies. Signals with broader signal content will not be differentiated correctly, particularly near the Nyquist frequency. The range of frequencies over which a gradient estimator is accurate will be referred to as the *derivative bandwidth*. For lack of a published definition, the derivative bandwidth will be defined as the

low-pass frequency span where the gradient estimator's magnitude is within $\pm 10\%$ of the ideal derivative's magnitude.

Examination of Figure 1.a promotes the idea that the 3-point central difference gradient estimator could be improved if the image is first filtered to eliminate the high-frequency content. The commonly suggested filter is a Gaussian smoothing filter based on sampling the analog Gaussian

$$h(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}}$$

where σ is 1.5–2 pixels and the spatial support of the digital filter is 9 or more taps. Figure 1.b shows the frequency response of a Gaussian filtered 3-point central difference gradient with $\sigma = 1.5$ and 9 taps. This filter is recommended in [10], but the frequency graphs show that it *reduces* the derivative bandwidth. The authors claim the smoothing removes aliasing effects and, thereby, improves optical flow accuracy. This claim may be more accurate for the temporal aliasing issue than the spatial aliasing issue (the authors do temporal as well as spatial smoothing of the data).

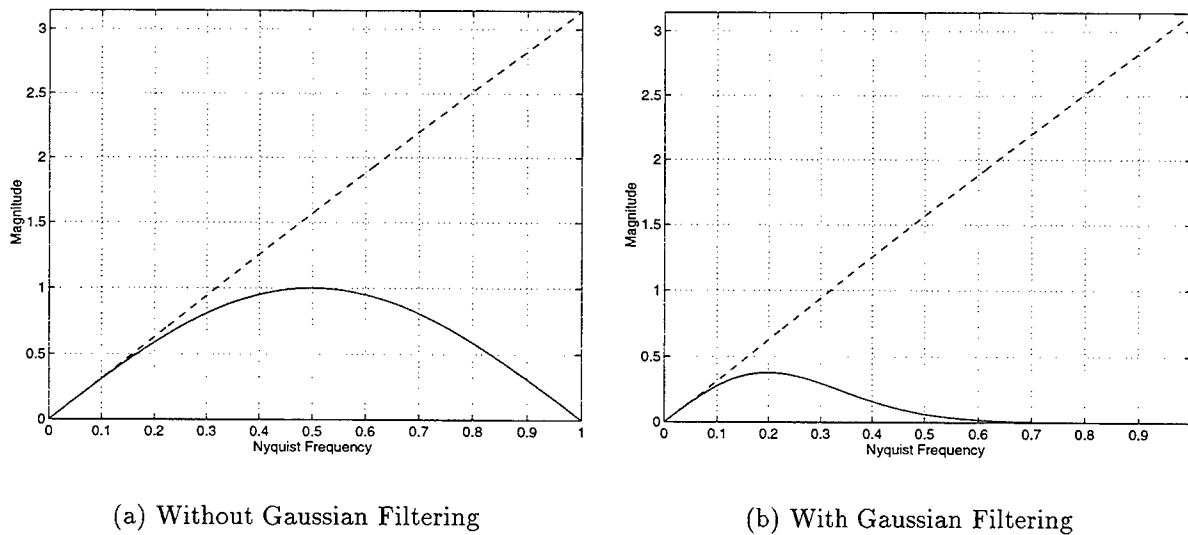


Figure 1: Magnitude of 3-point central difference gradient estimator with and without Gaussian filtering. Dashed line is an ideal derivative.

4.2.2 2-Point Causal Difference

The other common gradient estimator is the causal

$$h(z) = 1 - z^{-1} \tag{53}$$

with Fourier Transform

$$H(\omega) = 1 - e^{-j\omega} = j2e^{-j\omega/2} \sin(\omega/2)$$

It is used frequently for estimating temporal derivatives because it naturally arises in motion estimation algorithms such as the FITTS algorithm. Unlike (52), it is a causal filter and requires no processing delay. The phase of (53) is not the same as the ideal derivative, but the phase is linear and thus the group delay is constant with frequency. The magnitude, phase, and group delay are

$$|H(\omega)| = 2 \sin(\omega/2), \quad \angle H(\omega) = \frac{\pi}{2} - \frac{\omega}{2}, \quad -\frac{d\angle H(\omega)}{d\omega} = \frac{1}{2}$$

The magnitude is plotted in Figure 2 and has a derivative bandwidth of 0.25. This is much better than the 3-point central difference estimator. The 2-point causal difference (2PLD) also has a more extended high frequency response as well. However, the 2-point causal difference does have a half-pixel group delay.

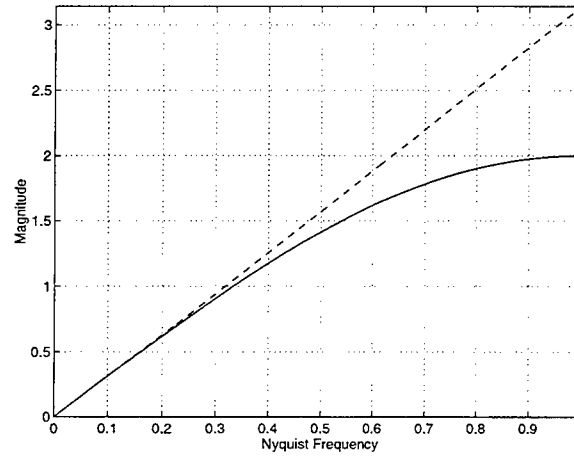


Figure 2: Magnitude of 2-point causal difference gradient estimator. Dashed line is an ideal derivative.

4.2.3 N -Point Central Differences

Let $y(x)$ be the analog signal to be differentiated. However, $y(x)$ is not available, only some sampled points of $y(x)$, $\{\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots\}$. Thus it is desired to find an N -length digital filter that can closely approximate the derivative of $y(t)$. This section demonstrates one method to find the filter weights of N -point central differences where N is an odd integer greater than one.

The technique is based on fitting a polynomial function

$$p(x) = a_{N-1}x^{N-1} + \dots + a_2x^2 + a_1x + a_0$$

to N of the $y(x)$ data points and then using the derivative of the polynomial to estimate $dy(x)/dx$. Proceed by way of example. Pick $N = 3$, and chose the data points centered around y_0 :

$$y_{-1} = p(-1), \quad y_0 = p(0), \quad y_1 = p(1)$$

With 3 data points, the polynomial is second order (otherwise the system of equations is over/under-determined). The system of equations to solve is

$$\begin{aligned} y_{-1} &= p(-1) = a_2 - a_1 + a_0 \\ y_0 &= p(0) = a_0 \\ y_1 &= p(1) = a_2 + a_1 + a_0 \end{aligned}$$

The solution is straightforward:

$$a_0 = y_0, \quad a_1 = \frac{y_1 - y_{-1}}{2}, \quad a_2 = \frac{y_1 + y_{-1}}{2}$$

The first derivative of $p(x)$ evaluated at $x = 0$ is

$$\hat{y}'_0 = \left. \frac{dp(x)}{dx} \right|_{x=0} = [2a_2x + a_1]_{x=0} = \frac{y_1 - y_{-1}}{2}$$

which is the 3-point central difference gradient estimator applied to the point in space $x = 0$. Polynomial fits can be made at other points of y_n using the 3 data points $\{y_{n-1}, y_n, y_{n+1}\}$ to find gradient estimates over all of $y(x)$. Performing such a procedure is equivalent to filtering (convolving) $\{\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots\}$ with the digital filter (52).

This technique can be extended to generate 5, 7, \dots , N -point central difference gradient estimators (N is odd) by increasing the number of data points used to generate the polynomial. However, the equations to be solved become increasingly more complicated. Fortunately, the polynomial interpolation problem has been well studied and a closed form solution for the polynomial is available, *Lagrange's Interpolation Formula* [11]. For the case at hand Lagrange's formula gives

$$p(x) = \sum_{k=-m}^m y_k \prod_{\substack{j=-m \\ j \neq k}}^m \left(\frac{x-j}{k-j} \right) \quad (54)$$

where $m = \frac{N-1}{2}$, and N is the number of data points. Filter weights are then computed by evaluating $p(x)'$ at $x = 0$. Appendix B lists a MATLABTM program that finds these weights. The output of this program is presented in Figure 3 for selected N .

4.3 N -Point Causal Differences

The polynomial fit technique can be used to find N -point causal gradient estimators as well. For the $N = 2$ case

$$y_{-1} = p(-1), \quad y_0 = p(0)$$

are the data points. The system of equations to be solved is

$$y_{-1} = -a_1 + a_0, \quad y_0 = a_0$$

with the obvious solution

$$a_0 = y_0, \quad a_1 = y_0 - y_{-1}$$

Then $p'(x)$ evaluated at $x = 0$ is

$$\left. \frac{dp(x)}{dx} \right|_{x=0} = y_0 - y_{-1}$$

which is the 2-point causal difference gradient estimator (53). Extended to 3, 4, 5, ..., N data points

$$p(x) = \sum_{k=-n+1}^0 y_k \prod_{\substack{j=-n+1 \\ j \neq k}}^0 \left(\frac{x-j}{k-j} \right) \quad (55)$$

Appendix C lists a MATLABTM program that finds the causal difference coefficients, and Figure 4 presents results for selected N .

4.4 Frequency Responses

Examination of the filter weights for N -point central difference gradient estimators shows that these filters are all anti-symmetric. That is,

$$h_N(z) = a_m z^m + a_{m-1} + \cdots + a_1 z^1 - a_1 z^{-1} + \cdots + a_{m-1} z^{-m+1} - a_m z^{-m} \quad (56)$$

where $m = \frac{N-1}{2}$. Substituting $z = e^{j\omega}$ shows that the frequency response is

$$H_N(\omega) = j2a_m \sin(m\omega) + j2a_{m-1} \sin([m-1]\omega) + \cdots + j2a_1 \sin(\omega) \quad (57)$$

Thus, all central difference gradient estimators have the desired phase for a derivative estimator (no group delay).

N	Filter Weights	Grad. Bandwidth
3	$\frac{1}{2} \times \{1, 0, -1\}$	0.125
5	$\frac{1}{12} \times \{-1, 8, 0, -8, 1\}$	0.223
7	$\frac{1}{60} \times \{1, -9, 45, 0, -45, 9, -1\}$	0.272
9	$\frac{1}{840} \times \{-3, 32, -168, 672, 0, -672, 168, -32, 3\}$	0.303
11	$\frac{1}{2520} \times \{2, -25, 150, -600, 2100, 0, -2100, 600, -150, 25, -2\}$	0.324

Figure 3: Central difference gradient weights.

Figure 5.a shows the magnitude response of central difference gradient estimators for $N \in \{3, 5, 7, \dots, 19\}$. As can be seen, increasing N improves the derivative bandwidth of the gradient estimators while maintaining similarly shaped magnitude responses. The responses are smooth and have no wild oscillations.

The frequency responses of causal difference gradient estimators do not have well behaved forms. Figure 5.b shows the magnitude response of central difference gradient estimators for $N \in \{2, 3, 4, 5, 6, 7\}$. As can be seen, increasing N does not improve the derivative bandwidth. Moreover, the high-frequency response of the gradient estimators becomes progressively more misbehaved. This is not very desirable behavior.

4.5 Open-Loop FITTS Algorithm

Having several different numerical derivative estimators available, the next step is to examine the performance of each in the FITTS algorithm. The first analysis is for an open-loop tracking system. The camera is stationary and a target travels through the FOV. The tracking algorithm measures the target's velocity.

4.5.1 Standard FITTS

Analyzing the SINE image in the FITTS algorithm is straightforward. Spatial gradients are computed using a 3PCD gradient estimator, while temporal gradients are computed with a 2PLD gradient estimator. These yield the sums

$$\sum \sum g_1^2 = \sin^2(2\pi f_1) \sum_{n_1} \cos^2(2\pi f_1[n_1 - V_1 t]) \sum_{x_2} \sin^2(2\pi f_2[x_2 - V_2 t]) \quad (58)$$

$$\sum \sum g_2^2 = \sin^2(2\pi f_2) \sum_{n_1} \sin^2(2\pi f_1[n_1 - V_1 t]) \sum_{n_2} \cos^2(2\pi f_2[n_2 - V_2 t]) \quad (59)$$

$$\sum \sum g_1 g_2 = \frac{1}{4} \sin(2\pi f_1) \sin(2\pi f_2) \sum_{n_1} \sin(4\pi f_1[n_1 - V_1 t]) \sum_{n_2} \sin(4\pi f_2[n_2 - V_2 t]) \quad (60)$$

where some manipulations with trigometric identities have been used to generate these forms. The sums in (60) are zero when the image frame contains an integer number of periods of the n_1 - and n_2 -direction sinusoids, and much smaller than the gradient sums (58) and (59). Thus, $\sum \sum g_1 g_2$ will be assumed to be zero. The FITTS algorithm in the n_1 -direction is then

$$d_1 \approx - \frac{\sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \Delta I[n_1, n_2] g_1[n_1, n_2]}{\sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} g_1^2[n_1, n_2]} \quad (61)$$

N	Filter Weights	Grad. Bandwidth
2	$\{1, -1\}$	0.250
3	$\frac{1}{2} \times \{3, -4, 1\}$	0.092
4	$\frac{1}{6} \times \{11, -18, 9, -2\}$	0.127
5	$\frac{1}{12} \times \{25, -48, 36, -16, 3\}$	0.178
6	$\frac{1}{60} \times \{127, -300, 300, -200, 75, 12\}$	0.225
7	$\frac{1}{60} \times \{147, -360, 450, -400, 225, -72, 10\}$	0.254

Figure 4: Causal difference gradient weights.

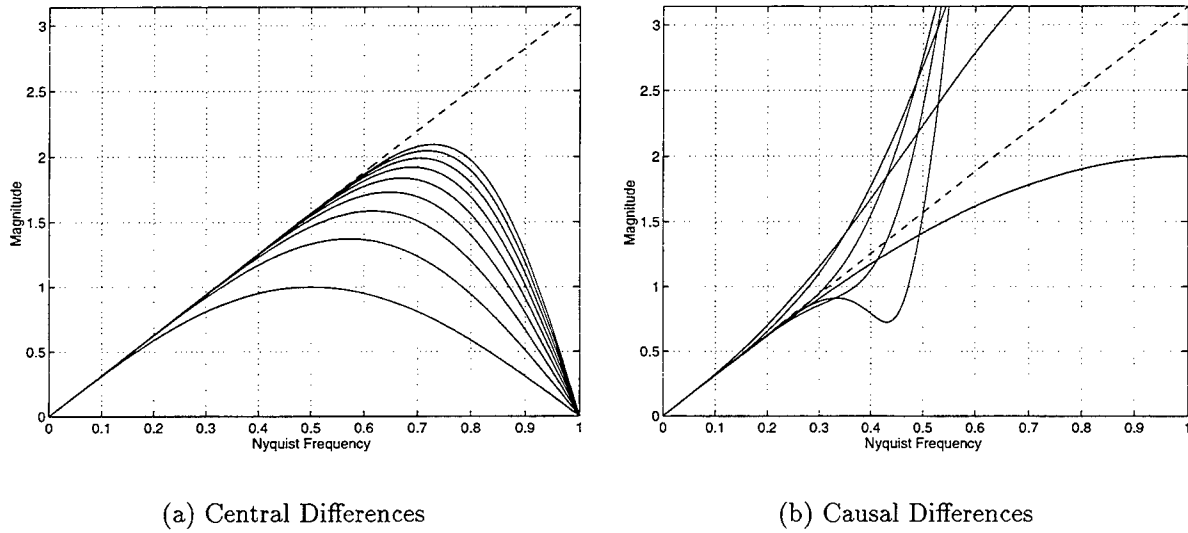


Figure 5: Magnitude of selected N -point difference gradient estimators. Dashed line is an ideal derivative.

The term ΔI is merely a 2PLD temporal derivative for which one computes

$$\begin{aligned}
\sum \sum \Delta I g_1 = & \sum_{n_1} \sum_{n_2} \sin(2\pi f_1) \cos(2\pi f_1[n_1 - V_1 t]) \sin(2\pi f_2[n_2 - V_2 t]) \times \\
& [\sin(2\pi f_1[n_1 - V_1 t]) \cos(2\pi f_2[n_2 - V_2 t]) - \\
& \cos(2\pi f_1 V_1) \cos(2\pi f_2 V_2) \sin(2\pi f_1[n_1 - V_1 t]) \sin(2\pi f_2[n_2 - V_2 t]) + \\
& \sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2) \cos(2\pi f_1[n_1 - V_1 t]) \sin(2\pi f_2[n_2 - V_2 t]) + \\
& \cos(2\pi f_1 V_1) \sin(2\pi f_2 V_2) \sin(2\pi f_1[n_1 - V_1 t]) \cos(2\pi f_2[n_2 - V_2 t]) - \\
& \sin(2\pi f_1 V_1) \sin(2\pi f_2 V_2) \cos(2\pi f_1[n_1 - V_1 t]) \cos(2\pi f_2[n_2 - V_2 t])] \quad (62)
\end{aligned}$$

Of the terms in (62), only the third ends up contributing to the summation. The other terms sum to zero, or nearly zero. Thus, $\sum \sum \Delta I g_1$ is effectively

$$\sum \sum \Delta I g_1 \approx \sin(2\pi f_1) \sum_{n_1} \cos^2(2\pi f_1[n_1 - V_1 t]) \sum_{n_2} \sin^2(2\pi f_2[n_2 - V_2 t])$$

Using this result, the FITTS equation in the n_1 -direction yields

$$d_1 \approx \frac{\sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2)}{\sin(2\pi f_1)} = \frac{\frac{1}{2} [\sin(2\pi[f_1 V_1 - f_2 V_2]) + \sin(2\pi[f_1 V_1 + f_2 V_2])]}{\sin(2\pi f_1)}$$

The computation of d_2 is unenlightenly similar due to the symmetry of the FITTS algorithm.

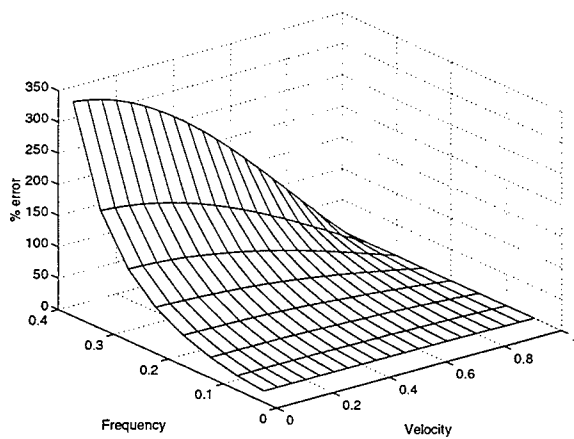
Figure 6.a shows the percentage error for the standard FITTS algorithm with $V_2 = 0$. Notice the peculiar trend: as velocity increases, the spatial bandwidth for accurate motion estimation also increases. This result is not intuitive, but is easily explained. The absolute error is small for low velocities, but when divided by the small velocity to be estimated, gives a large percentage of error. Also note the selection of $V_2 = 0$. The presence of velocity in the n_2 -direction affects the effort to compute velocity in the n_1 -direction. This is even with the cross-coupling terms in the FITTS algorithm and using a spatially separable image.

Since ΔI is a temporal gradient, it may be possible to improve the FITTS algorithm by using a 3PCD for the temporal gradient. Then,

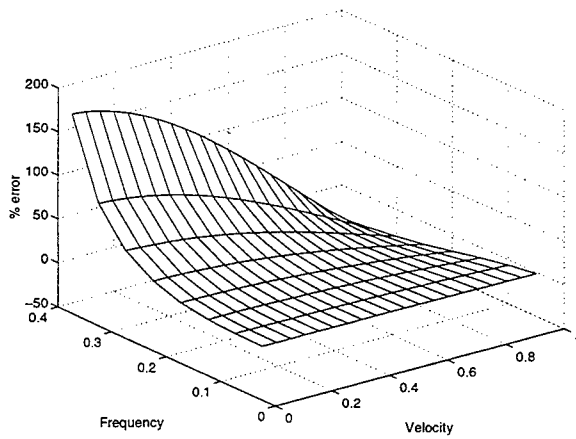
$$\begin{aligned}
\sum \sum \Delta I g_1 = & \sin(2\pi f_1) \sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2) \sum_{n_1} \cos^2(2\pi f_1[n_1 - V_1 t]) \sum_{n_2} \sin^2(2\pi f_2[n_2 - V_2 t]) + \\
& \sin(2\pi f_1) \cos(2\pi f_1 V_1) \sin(2\pi f_2 V_2) \sum_{n_1} \sin(2\pi f_1[n_1 - V_1 t]) \cos(2\pi f_1[n_1 - V_1 t]) \\
& \sum_{n_2} \sin(2\pi f_2[n_2 - V_2 t]) \cos(2\pi f_2[n_2 - V_2 t])
\end{aligned}$$

The second term of $\sum \sum \Delta I g_1$ is nearly zero. Thus,

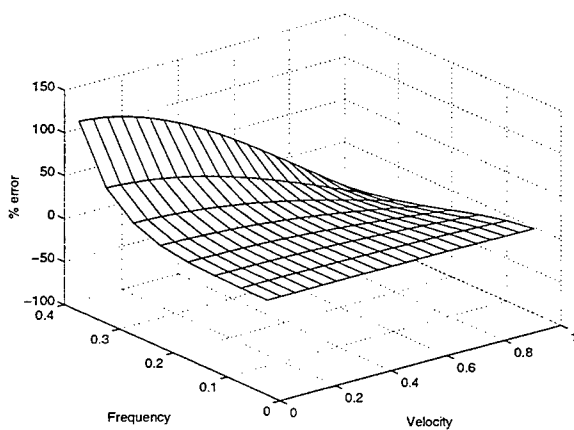
$$d_1 \approx \frac{\sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2)}{\sin(2\pi f_1)} = \frac{\frac{1}{2} [\sin(2\pi[f_1 V_1 - f_2 V_2]) + \sin(2\pi[f_1 V_1 + f_2 V_2])]}{\sin(2\pi f_1)}$$



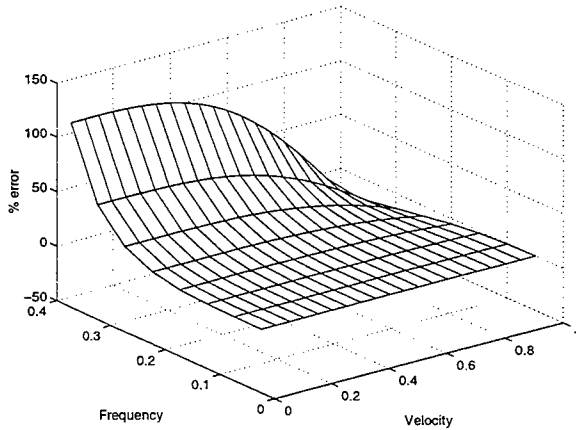
(a) 3PCD spatial gradient estimator with 2PLD temporal gradient estimator



(b) 5PCD spatial gradient estimator with 2PLD temporal gradient estimator

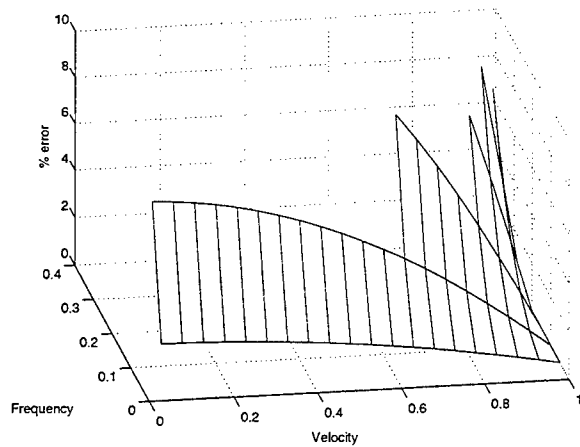


(c) 7PCD spatial gradient estimator with 2PLD temporal gradient estimator

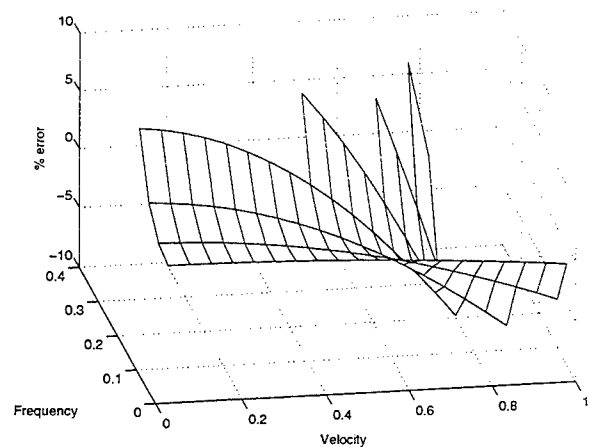


(d) 7PCD spatial gradient estimators with 5PCD temporal gradient estimator

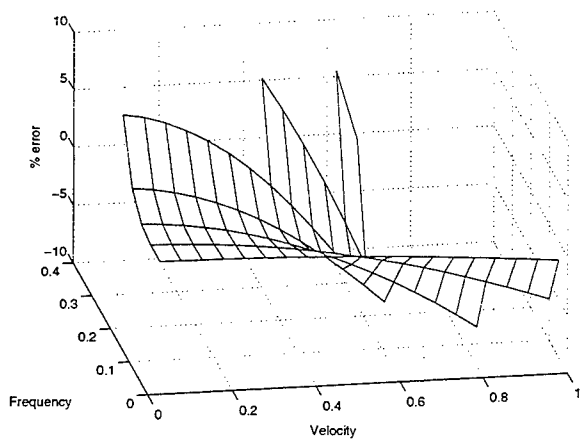
Figure 6: Performance of FITTS with SINE image.



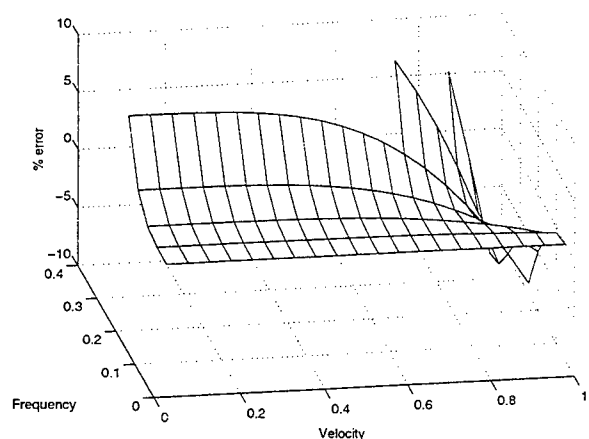
(a) 3PCD spatial gradient estimator & 2PLD temporal gradient estimator



(b) 5PCD spatial gradient estimator with 2PLD temporal gradient estimator



(c) 7PCD spatial gradient estimator with 2PLD temporal gradient estimator



(d) 7PCD spatial gradient estimators with 5PCD temporal gradient estimator

Figure 7: Expanded views of performance of FITTS with SINE image.

which is an identical result to the 2PLD gradient results. Although this suggests improving the temporal derivative bandwidth does not help, this is not true, as will be shown by utilizing a 5PCD for the temporal gradient.

Further examination of FITTS performance can be made by repeating the previous gradient calculations with various spatial and temporal gradient filters. For a 5PCD spatial gradient and a 3PCD temporal gradient,

$$\sum \sum g_1^2 = \frac{1}{36} (8 \sin(2\pi f_1) - \sin(4\pi f_1)) \sum_{n_1} \cos^2(2\pi f_1 [n_1 - V_1 t]) \sum_{n_2} \sin^2(2\pi f_2 [n_2 - V_2 t])$$

and

$$\begin{aligned} \sum \sum \Delta I g_1 = & \frac{1}{6} (8 \sin(2\pi f_1) - \sin(4\pi f_1)) \left[\sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2) \sum_{n_1} \cos^2(2\pi f_1 [n_1 - V_1 t]) \right. \\ & \sum_{n_2} \sin^2(2\pi f_2 [n_2 - V_2 t]) + \cos(2\pi f_1 V_1) \sin(2\pi f_2 V_2) \sum_{n_1} \sin(2\pi f_1 [n_1 - V_1 t]) \\ & \left. \cos(2\pi f_1 [n_1 - V_1 t]) \sum_{n_2} \sin(2\pi f_2 [n_2 - V_2 t]) \cos(2\pi f_2 [n_2 - V_2 t]) \right] \end{aligned}$$

giving

$$d_1 \approx \frac{\sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2)}{\frac{1}{6} [8 \sin(2\pi f_1) - 2 \sin(4\pi f_1)]}$$

The performance of this gradient combination is illustrated in Figure 6.b. Continuing the analysis, if the spatial gradient is a 7PCD then

$$d_1 \approx \frac{\sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2)}{\frac{1}{30} [45 \sin(2\pi f_1) - 9 \sin(4\pi f_1) + \sin(6\pi f_1)]}$$

Last, let the temporal gradient be a 5PCD with a 7PCD spatial gradient

$$d_1 \approx \frac{\frac{1}{6} [8 \sin(2\pi f_1 V_1) \cos(2\pi f_2 V_2) - \sin(4\pi f_1 V_1) \cos(4\pi f_2 V_2)]}{\frac{1}{30} [45 \sin(2\pi f_1) - 9 \sin(4\pi f_1) + \sin(6\pi f_1)]}$$

A clear trend is forming for the SINE image. **The estimate of the velocity is the temporal derivative divided by the spatial derivative.** However, it must be emphasized that this analysis of the FITTS algorithm is **not** a transfer function analysis. The FITTS algorithm is not linear in intensity, and thus what can be said about the algorithm for general images is not a direct extension of the SINE image results.

Figures 6.c-d show the performance of the last two versions of FITTS. Figures 7.a-d show expanded views of the previous results which show only the region of the FITTS algorithm response that has less than 10% error. These plots show the open-loop performance improvements in the FITTS algorithm with increasing spatial and temporal gradient size.

4.5.2 Verification of SINE Performance

The SINE image was employed in an open loop FITTS simulation to verify the calculations of performance. This demonstrated an important point about the FITTS computation. The gradient estimate at the edges of the image requires data that is unknown. Consequently, accurate gradient estimation at the edges is not possible. In initial simulations, these outside pixels were assumed to be zero. This resulted in about a 10% difference between simulated FITTS tracking performance and the predicted performance. Therefore, it is recommended that the FITTS algorithm be implemented with the summation indexes $n_1 = 2, \dots, (N_1 - 1)$ and $n_2 = 2, \dots, (N_2 - 1)$ instead of the full summation ranges for 3PCD spatial gradients. This corrected the simulation results so that they matched the expected values from the computations. The problem gets worse with higher order spatial gradients. A 5PCD spatial gradient requires pixel values outside the image boundaries not only on the edge, but for the pixels next to the edge as well. The solution is to change the FITTS summation indexes to $n_1 = 3, \dots, (N_1 - 2)$ and $n_2 = 3, \dots, (N_2 - 2)$. Analogous situations and solutions hold for N -point central difference spatial gradients.

4.6 Closed Loop FITTS

The analysis up to this point has been for the FITTS algorithm without the use of the MAP drift correction terms (8-9) or within a closed-loop simulation. As has been shown, for the open-loop FITTS algorithm the gradients are very significant to performance. Does this carry over to the closed-loop algorithm? This question is more difficult to answer, as an analysis is highly non-linear and intractable. Therefore, the question will be answered by performing simulations of the FITTS algorithm with the drift removal terms. Again, a sinusoidal image will be used. In this case, the accumulated displacements

$$\overline{d_1[k]} = \sum_{i=1}^k d_1[i], \quad \overline{d_2[k]} = \sum_{i=1}^k d_2[i] \quad (63)$$

are used to move the camera to maintain tracking. Thus, the received camera image will be

$$I[n_1, n_2; k] = \sin(2\pi f_1[n_1 - V_1 k + \overline{d_1[k-1]}]) \sin(2\pi f_2[n_2 - V_2 k + \overline{d_2[k-1]}]) \quad (64)$$

which will be referred to as SINE2.

Several values and combinations of f_1 and f_2 were examined in conjunction with various values of V_1 and V_2 . Performance was determined by examining the tracking errors

$$e_1[k] = \overline{d_1[k]} - V_1 k, \quad e_2[k] = \overline{d_2[k]} - V_2 k \quad (65)$$

for transient performance and instability. The table in Figure 8 lists a few sets of experiments that provide information. For the experiments, the MAP update weight ρ in (7) was set to a value of 0.1 [3]. The value of ρ was found to influence tracking primarily in determining the time constant for the error's transient

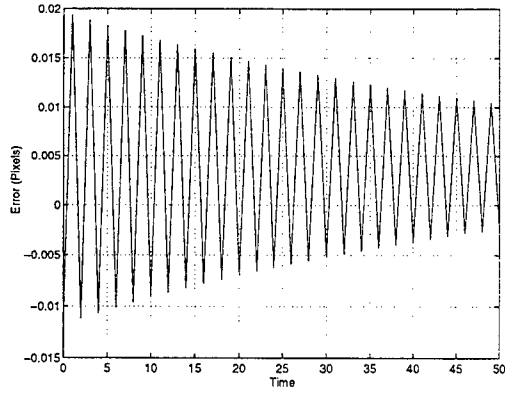
response. It did not affect stability significantly. Other simulation parameters are listed in the table along with the largest velocity that could be tracked (estimated to within a tenth) under two conditions. The first condition was a target with velocity in only the n_1 -direction, that is, $V_2 = 0$, and the second condition was equal target velocities in each direction, that is, $V_1 = V_2$.

Clearly, at low spatial frequencies the range of trackable velocities is very great. Also, there is a trend in the trackable target velocities: target motion in both directions has half the tracking range as motion in one direction only. Notice that the table stops at velocities of 0.31 and up, and does not show any spatial frequency combinations with unequal frequencies. This is because it was found that a spatial frequency of 0.3 in *either* spatial direction was the tracking "red line" for the SINE2 image. Figure 9 shows the performance near the "red line" for f_1 values of 0.300 and 0.303 when $f_2 = 0.05$ and $V_1 = V_2 = 0.01$.

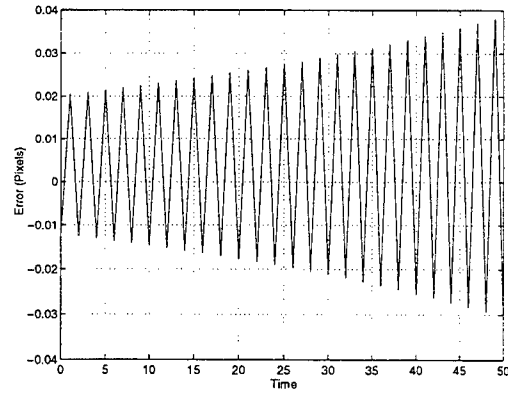
The effect of a gradient estimator's derivative bandwidth was less significant than in the open-loop FITTS algorithm. Primarily, using a superior gradient made a small improvement in trackable velocities. Transient behavior was not improved except near the tracking "red line." An example is illustrated in Figure 10 for the case $f_1 = f_2 = 0.25$ and $V_1 = V_2 = 0.5$.

f_1	f_2	Gradient	Maximum Velocity Tracked
0.05	0.05	3PCD	$V_1 = 4.5$ ($V_2 = 0$), $V_1 = V_2 = 2.4$
0.05	0.05	5PCD	$V_1 = 4.5$ ($V_2 = 0$), $V_1 = V_2 = 2.4$
0.05	0.05	7PCD	$V_1 = 4.6$ ($V_2 = 0$), $V_1 = V_2 = 2.4$
0.15	0.15	3PCD	$V_1 = 1.7$ ($V_2 = 0$), $V_1 = V_2 = 0.9$
0.15	0.15	5PCD	$V_1 = 1.8$ ($V_2 = 0$), $V_1 = V_2 = 0.9$
0.15	0.15	7PCD	$V_1 = 1.8$ ($V_2 = 0$), $V_1 = V_2 = 0.9$
0.25	0.25	3PCD	$V_1 = 1.0$ ($V_2 = 0$), $V_1 = V_2 = 0.5$
0.25	0.25	5PCD	$V_1 = 1.4$ ($V_2 = 0$), $V_1 = V_2 = 0.7$
0.25	0.25	7PCD	$V_1 = 1.4$ ($V_2 = 0$), $V_1 = V_2 = 0.7$
0.30	0.30	3PCD	$V_1 = 0.05$ ($V_2 = 0$), $V_1 = V_2 = 0.03$
0.30	0.30	5PCD	$V_1 = 0.1$ ($V_2 = 0$), $V_1 = V_2 = 0.05$
0.30	0.30	7PCD	$V_1 = 0.1$ ($V_2 = 0$), $V_1 = V_2 = 0.05$
>0.31	>0.31	3PCD	Tracking fails for all velocities.
>0.31	>0.31	5PCD	
>0.31	>0.31	7PCD	

Figure 8: Results of SINE2 in the complete FITTS algorithm.

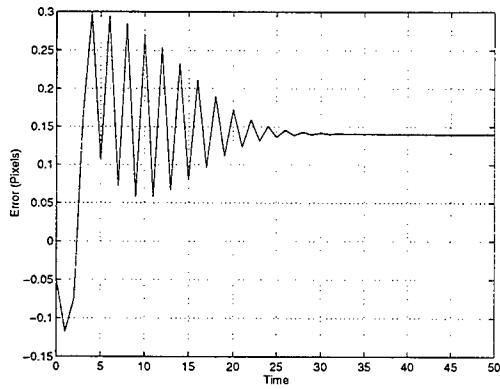


(a) $f_1 = 0.300$

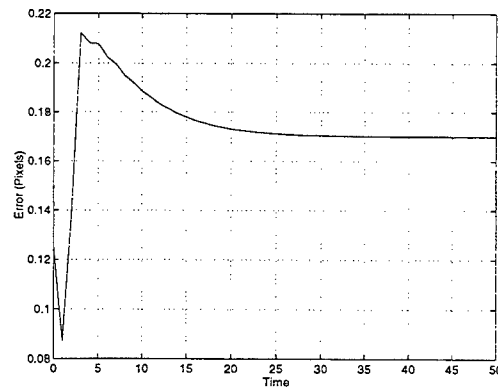


(b) $f_2 = 0.303$

Figure 9: Comparison of SINE2 performance near the $f_1 = 0.3$ boundary. $f_2 = 0.05$, $V_1 = V_2 = 0.01$, and the gradient is 3PCD.



(a) 3PCD



(b) 7PCD

Figure 10: Comparison of SINE2 performance using the 3PCD and 7PCD gradient estimation schemes. $f_1 = f_2 = 0.25$ and $V_1 = V_2 = 0.5$.

5 Missile Simulation Experiments

The previous sections have developed two possible improvements in FITTS type missile tracking. First, it was shown that gradients with greater derivative bandwidth can improve open-loop velocity estimation, and possibly improve closed-loop tracking. Second, an affine motion model FITTS algorithm was developed that allows greater degrees of freedom in the missile's movement. In this section, these two potential improvements will be examined for their efficacy.

5.1 Missile Sequence

Two sequences of missile images were available for the simulations. These missile images were synthetically generated by an independent computer simulation and supplied by Phillips laboratory. The first sequence will be referred to as the OLD missile data, and the second as the NEW missile data. Each sequence contains 101 frames of 64×64 missile image data that represent a 32μ -radian by 32μ -radian tracker field of view. The missile is being imaged through a 2 meter aperture at a range of 200 km. In the OLD missile sequence the missile travels with a velocity of 900 m/s. Illumination of the missile is assumed to be a uniform laser at the missile. The NEW missile sequence is based on a velocity of 970 m/s, but much more important, the missile is illuminated with a wide beam laser from the tracker of 1000 watts power.

In addition to the missile images, an additional set of images was synthesized to go along with the OLD and NEW missile sequences. This was of a High Energy Laser (HEL) propagated to the missile through the 2 meter aperture. The propagation time to the missile is long enough that the HEL arrival point is affected by the missile's velocity. Consequently, HEL aiming is compensated by pointing the HEL ahead of the aim point approximately 7 pixels. The propagation of the light for each missile and HEL sequences was performed simultaneously to simulate the actual conditions that would be present during operation of a tracking system. It is worth noting that by storing the individual light propagations in separate sequences, the atmosphere is assumed to be linear so that the principle of superposition applies. The atmosphere through which the light is propagating is moderately turbulent which causes a characteristic blurring and shifting of each image sequence. The HEL is a different wavelength than the missile illumination and therefore suffers slightly different distortions in passing through the atmosphere than the missile data.

The 101 missile images are denoted $I[n_1, n_2; 0]$ through $I[n_1, n_2; 100]$. The n_1 -direction is the vertical axis ("y-axis") and the n_2 -direction is the horizontal ("x-axis"). (This corresponds with the co-ordinate system used by MATLABTM to display image data.) The first image of the OLD and NEW sequences is shown in Figure 11 with the co-ordinate axis labeled.

5.2 Algorithm Implementation Details

Several issues concerning the implementation and evaluation of the experiments are worth noting.

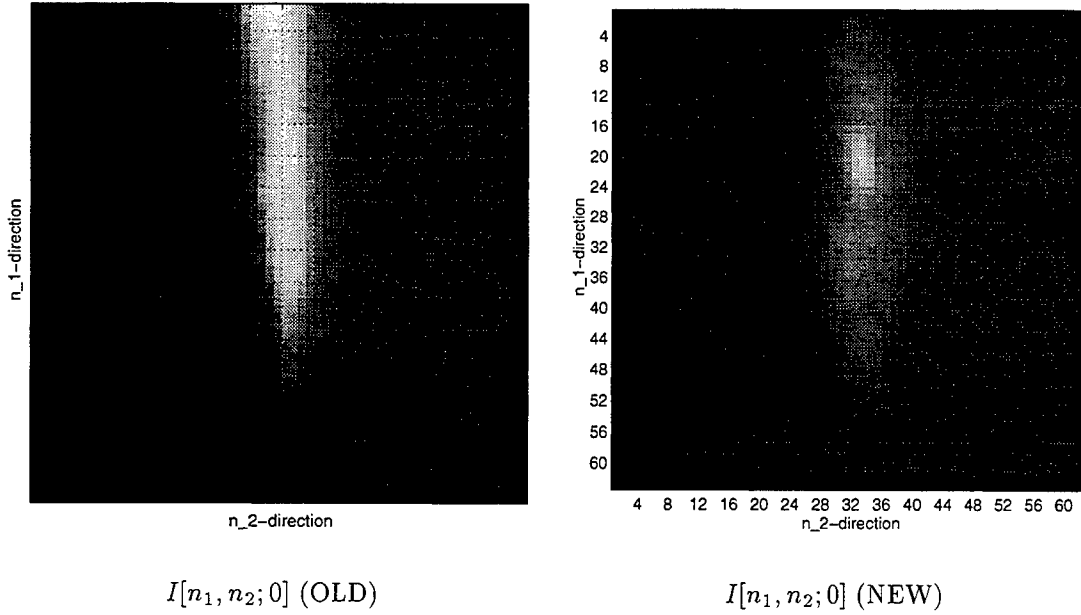


Figure 11: Sample missile images for the NEW and OLD sequences.

5.2.1 Modified Drift Correction

As has been previously reported [3], the FITTS algorithm has difficulty tracking the missile data due to the atmospheric distortions. At first the FITTS drift correction equation (9) was modified to

$$d_1[k] = d_{1corr}[k] + d_{1map}[k] + (\text{cen}_{1map}[k] - \text{cen}_{1map}[0]) \quad (66)$$

where $\text{cen}_{1map}[k]$ is the n_1 -direction centroid of $MAP[n_1, n_2, k]$. (An analogous equation applies for the n_2 -direction as well.) Adding this MAP centroid correction resulted in fair rejection of the atmospheric noise disturbances, but not complete rejection of them. When the value of ρ was increased to 0.2 or more the tracker drift was noticeable and unacceptable. Thus, the drift correction was changed to

$$d_1[k] = d_{1corr}[k] + (\text{cen}_{1map}[k] - \text{cen}_{1map}[0]) \quad (67)$$

which only has the MAP centroid compensation. This was found to work well in eliminating the drift problem for useful values of ρ .

5.2.2 Image Interpolation

To simulate the affects of a closed-loop tracker, the images need to be shifted to simulate camera movement. The continuous image frame is not available which makes it impossible to compute images for non-integer pixel motions. However, it is necessary to do so because the FITTS algorithm tracking output is non-integer. The problem was overcome by using *bilinear interpolation* to interpolate non-integer co-ordinate

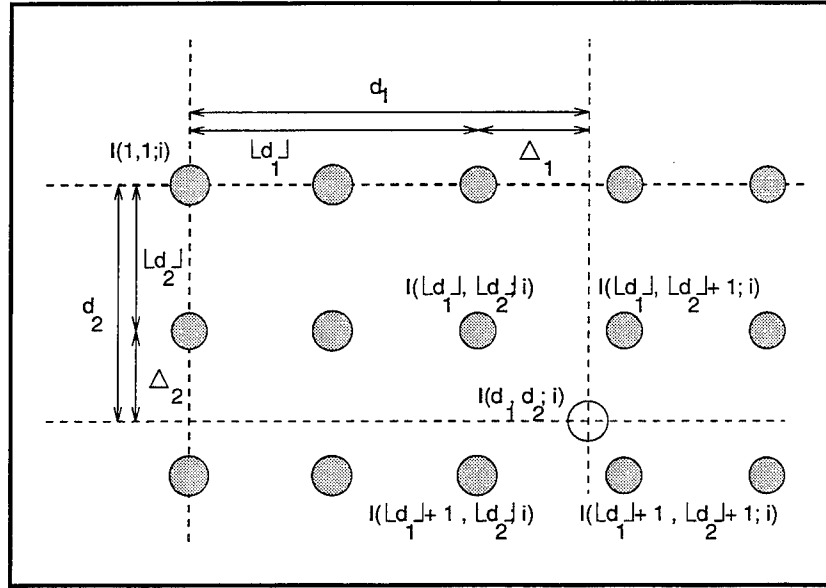


Figure 12: Illustration of bilinear interpolation procedure.

pixel intensities [12]

$$\begin{aligned}
 I(n_1, n_2; i) = & I(\lfloor n_1 \rfloor, \lfloor n_2 \rfloor; i)(1 - \Delta_1)(1 - \Delta_2) + I(\lfloor n_1 \rfloor + 1, \lfloor n_2 \rfloor; i)\Delta_1(1 - \Delta_2) + \\
 & I(\lfloor n_1 \rfloor, \lfloor n_2 \rfloor + 1; i)(1 - \Delta_1)\Delta_2 + I(\lfloor n_1 \rfloor + 1, \lfloor n_2 \rfloor + 1; i)\Delta_1\Delta_2
 \end{aligned} \tag{68}$$

where $\Delta_1 = n_1 - \lfloor n_1 \rfloor$ and $\Delta_2 = n_2 - \lfloor n_2 \rfloor$. The operator $\lfloor \cdot \rfloor$ represents the floor operator (closest integer to the operator that is less than or equal to it). If pixels outside the image frame boundaries were needed, their values were interpolated to be the value of the nearest boundary pixel. Figure 12 illustrates (68).

5.2.3 Initializing the MAP

The MAP has to be initialized for the FITTS algorithm. Two techniques were used in the simulations:

- $MAP[k] = I[n_1, n_2; k]$, that is, the MAP starts as the first image in the sequence.
- The missile image sequence $I[n_1, n_2; k]$ was taken in reverse order and put through the MAP filter (7). Before each image was filtered in (7), it was shifted using bilinear interpolation such that its centroid was identical to the centroid of the first image, $I[n_1, n_2; 0]$. The final image will be denoted $\overline{MAP}[n_1, n_2; 0]$.

The first technique simulates tracking from a cold start, while the second simulates tracking after the tracker has been running for a while.

5.2.4 FITTS Performance Evaluation

The overall goal of the tracker is to provide tilt information which will allow the HEL to be pointed accurately at the missile aim point. To quantify the accuracy of using tilts from a calculation to estimate an unknown tilt, the correlation coefficient between the tilts will be computed. This calculation is

$$\rho(c[k], d[k]) = \frac{\sum_{i=0}^{100} (c[i] - \bar{c})(d[i] - \bar{d})}{\sqrt{\sum_{i=0}^{100} (c[i] - \bar{c})^2 \sum_{i=0}^{100} (d[i] - \bar{d})^2}} \quad (69)$$

where $c[i], d[i]$ are any two tilts (measured or actual) and \bar{c}, \bar{d} are the average values of these tilts, respectively. The FITTS algorithm output was integrated using (63) to convert the velocity information into pixel position so that it could be compared with the centroid of the HEL.

The correlation parameter indicates how much output error variance can be expected by substituting one correlated signal in place of the other. For the missile tracker the desire is to use the missile tilt information to point the HEL correctly. In particular the error variance from the substitution can be found exactly with

$$\text{Var} \left\{ c[k] - \frac{\sigma_c}{\sigma_d} d[k] \right\} = [1 - \rho(c[k], d[k])^2] \sigma_c^2 \quad (70)$$

where σ_c^2 is the variance of $c[k]$ and σ_d^2 is the variance of $d[k]$. Thus, a correlation of 0.5 means 25% of the (normalized) power is similar between the two processes. A correlation of 0.7 brings this up to 49%, and a correlation of 0.9 represents only 81% similarity. This square law behavior demonstrates that very high correlations are desirable for good HEL pointing.

5.2.5 A-FITTS Performance Evaluation

The affine FITTS motion does not immediately lend itself to an easy centroid comparisons like the standard FITTS algorithm. Moreover, there are not currently plans for the tracker camera to adjust itself in an affine manner for missile tracking. Consequently, the optical-flow-to-centroid conversion equation developed in section 2 was used to convert affine optical flow into a tilt measurement. The procedures in 5.2.4 were then applied to calculate correlations.

5.3 Simulation Results

A simple set of experiments was run on both the NEW and OLD data sets. The parameters for an results of these experiments are tabulated in Figures 13 and 14. First, ρ was varied for the MAP filtering to select a good value for other experiments. This demonstrated contradictory biases. For the NEW data set, if the MAP was initialized with the first image of the sequence, smaller ρ were better, while if MAP started as the average of the images a value of 0.10 was best. For the OLD data set, increasing ρ always improved the results. Since the NEW data set is thought to be the more realistic missile scenario, ρ was set to 0.05 for subsequent simulations.

Next, different gradients were tried in the FITTS and A-FITTS algorithms. The results were inconclusive. Although small variations in the correlations were found, the differences were quite small and hardly statistically significant. The only exception is the A-FITTS algorithm using the OLD data set and an average initial MAP. For those experiments the A-FITTS correlation coefficient was about 0.05 greater than for the FITTS under similar conditions. If anything, it can be concluded that the improvements seen in the FITTS algorithm for analytical data do not help nor hinder missile tracking. It is also worth remembering that the affine FITTS algorithm output is being converted to tilts by the optical-flow-to-centroid conversion. There is still the possibility of performance improvements if the affine motion can be utilized effectively.

The last comment is to note the n_1 -direction correlation for the NEW data is half that of the OLD data. One possible explanation for this is the size of the illumination laser spot. This spot does not fully cover the missile contained in the field of view creating a "false" missile edge at the top. Consequently, any jitter of the illumination laser as it propagates is going to appear as movement of this "false" top edge. Since the FITTS algorithms and many other optical flow algorithms use edge information (this is where the largest gradients are) for tracking, there has to be a detrimental affect. For the case at hand, it is entirely possible that most of the performance degradation between the NEW and OLD data set is from this effect.

Algorithm	$MAP[n_1, n_2; 0]$	ρ	Spatial Gradient	HEL Centroid Correlations	
				n_1 -direction	n_2 -direction
FITTS	$I[n_1, n_2; 0]$	0.05	3PCD	0.4530	-0.9414
FITTS	$I[n_1, n_2; 0]$	0.10	3PCD	0.4258	-0.9432
FITTS	$I[n_1, n_2; 0]$	0.20	3PCD	0.4132	-0.9503
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	3PCD	0.3647	-0.9395
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.10	3PCD	0.3435	-0.9417
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.20	3PCD	0.3704	-0.9498
FITTS	$I[n_1, n_2; 0]$	0.05	5PCD	0.4500	-0.9412
FITTS	$I[n_1, n_2; 0]$	0.05	7PCD	0.4487	-0.9412
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	5PCD	0.3631	-0.9393
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	7PCD	0.3669	-0.9393
A-FITTS	$I[n_1, n_2; 0]$	0.05	3PCD	0.4291	-0.9387
A-FITTS	$I[n_1, n_2; 0]$	0.05	5PCD	0.4336	-0.9389
A-FITTS	$I[n_1, n_2; 0]$	0.05	7PCD	0.4404	-0.9391
A-FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	3PCD	0.3681	-0.9391
A-FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	5PCD	0.3725	-0.9393
A-FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	7PCD	0.3805	-0.9396

Figure 13: Performance for NEW missile data in terms of correlation coefficients.

Algorithm	$MAP[n_1, n_2; 0]$	ρ	Spatial Gradient	HEL Centroid Correlations	
				n_1 -direction	n_2 -direction
FITTS	$I[n_1, n_2; 0]$	0.05	3PCD	-0.7404	-0.8255
FITTS	$I[n_1, n_2; 0]$	0.10	3PCD	-0.8085	-0.8139
FITTS	$I[n_1, n_2; 0]$	0.20	3PCD	-0.8116	-0.8073
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	3PCD	-0.6949	-0.8343
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.10	3PCD	-0.7270	-0.8205
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.20	3PCD	-0.7885	-0.8092
FITTS	$I[n_1, n_2; 0]$	0.05	5PCD	-0.7539	-0.8354
FITTS	$I[n_1, n_2; 0]$	0.05	7PCD	-0.7632	-0.8455
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	5PCD	-0.6967	-0.8441
FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	7PCD	-0.6997	-0.8537
A-FITTS	$I[n_1, n_2; 0]$	0.05	3PCD	-0.7982	-0.7832
A-FITTS	$I[n_1, n_2; 0]$	0.05	5PCD	-0.7956	-0.7902
A-FITTS	$I[n_1, n_2; 0]$	0.05	7PCD	-0.7883	-0.7964
A-FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	3PCD	-0.7632	-0.7812
A-FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	5PCD	-0.7595	-0.7881
A-FITTS	$\overline{MAP}[n_1, n_2; 0]$	0.05	7PCD	-0.7547	-0.7941

Figure 14: Performance for OLD missile data in terms of correlation coefficients.

6 Conclusions and Recommendations

The FITTS algorithm is a special case of optical flow algorithms. This view point allows the FITTS algorithm to be examined analytically for potential areas of improvement in the missile tracking application. Two potential improvements were identified—using gradients with greater derivative bandwidth and allowing the motion to be affine—and further investigated. The results of this research have shown that theoretical improvements from gradients with greater derivative bandwidths do not appear to carry over to missile tracking. A small improvement in correlation was note for the OLD missile data set with the affine FITTS algorithm, so affine motion tracking *may* be able improve missile tracking. To make this determination it will be necessary to determine if the small illumination laser spot size used in the NEW data set was the cause of poor performance. Also, the affine motion model allows greater motion freedom than tilts, and this information may be useful to tracking.

The analysis of the FITTS algorithm drift correction has demonstrated that it may be fundamentally unstable. Previous simulations [3] have agreed with the current simulations which show that the FITTS drift corrections are completely ineffective with the missile data. One possible reason may be the nature of the missile data itself. The tilt of the missile image in the field of view is from two sources:

1. “Apparent” movement of the missile caused by atmospheric imaging distortions (scintillation noise).
2. “Real” missile movement caused by the missile’s velocity.

This can be mathematically approximated³ as

$$r(\mathbf{x};t) = m(\mathbf{x};t) [h(\mathbf{x};t) \star\star s(\mathbf{x} - t\mathbf{d}_m)] + w(\mathbf{x};t) \quad (71)$$

which is illustrated in Figure 15. The signal $s(\mathbf{x};t)$ represents the received missile images. These images are contaminated by the additive Gaussian white noise $w(\mathbf{x};t)$. The undistorted missile image, $s(\mathbf{x})$, is blurred and tilted by the *spatially varying* atmospheric distortion $h(\mathbf{x})$ through an operation $\star\star$ (the operation $\star\star$ is similar to a 2-dimensional convolution, but it is spatially invariant). The mask, $m(\mathbf{x};t)$ introduces a multiplicative attenuation noise onto the missile image. During the image formation process, the missile is moving through space with a velocity \mathbf{d}_m . The tracker need not distinguish between the true missile velocity \mathbf{d}_m or the tilts caused by the blurring distortion $h(\mathbf{x};t)$. In fact, discrimination between the tilt from the atmosphere and the missile velocity undesirable. This is because to hit the missile, the HEL must be pointed at the imaging system’s missile location, not the true missile location. The reason is the HEL beam passes through nearly the same atmosphere as the missile illumination return and, therefore, suffers similar atmospheric distortions. The difficulty is the *combined* effect of the atmospheric distortions and the

³This mathematical approximation is not necessarily a valid model of the image formation process. It is *ad hoc* and intended only to demonstrate the *observed* nature of the image data.

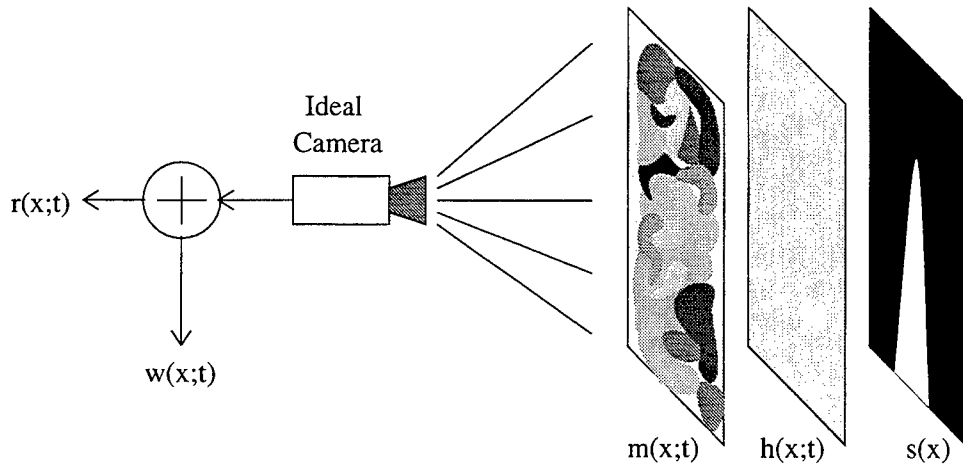


Figure 15: Suggested Image formation model.

missile's velocity. When the missile is properly tracked the camera will be shifted so that the field of view keeps the missile centered. Then received images can be modeled as

$$r(\mathbf{x}; t) = m(\mathbf{x} - t\mathbf{d}_m; t) [h(\mathbf{x}; t) \star \star s(\mathbf{x})] + w(\mathbf{x}; t) \quad (72)$$

The important feature in (72) is that under tracking lock, the multiplicative noise $m(\mathbf{x}; t)$ will translate across the blurred missile image. For the missile data observed to date, this noise is highly structured and temporally slowly varying. The structure of the noise has significant gradients everywhere while the missile image only has gradient content near the missile's edges. Consequently, typical optical flow algorithms identify the movement of the distortion phenomenon in preference or in conjunction with the tilt arising from the missile blur $h(\mathbf{x}; t)$ [3]. It may be necessary to formulate a new optical flow regime based on (72) as the imaging system model.

References

- [1] J.M. Fitts, "Correlation Tracking Via Optimum Weighting Functions," Report No. P73-240 from Hughes, Culver City, CA, 1973.
- [2] J.M. Fitts, "Optimal Weighting Function Correlation Tracker", Hughes Aircraft Co. Internal Report.
- [3] J.I. Lipp, "Estimation of Tilts of Extended Images in the Presence of Atmospheric Disturbances Using Optical Flow Algorithms," Final Report for Air Force Office of Scientific Research Summer Research Program, August 1994, pp. 8-1 to 8-20.

- [4] D.J. Fleet, and K. Langley, "Recursive Filters for Optical Flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 61-67, Jan. 1995.
- [5] M.R. Luetttgen, W.C. Karl, and A.S. Willsky, "Efficient Multiscale Regularization with Applications to the Computation of Optical Flow," *IEEE Transactions on Image Processing*, Vol. 3, No.1, January 1994, pp. 41-63.
- [6] B.K. Horn and B.G. Schunk, "Determining Optical Flow," *Artificial Intelligence*, Vol. 17, 1981, pp. 185-203.
- [7] D. Brunson, G. Grockett, and J.F. Riker, "Alternate Derivation of Correlation Track Algorithm," RDA (LOGICON) internal correspondence, March 1991.
- [8] Van Trees, H.L., *Detection, Estimation, and Modulation Theory - Part 1*. New York: Wiley, 1968.
- [9] J.I. Lipp, "Notes on Fitts' Correlation Tracker," Internal Report to Sal Cusamano, Philips Laboratory. September 1994.
- [10] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Performance of Optical Flow Techniques," *International Journal of Computer Vision*, 12:1, pp. 43-77, 1994.
- [11] Horn, R.A., and Johnson, C.R. *Matrix Analysis*. New York: Cambridge, 1990.
- [12] W.K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.
- [13] P.A. Nelson and S.J. Elliot, *Active Control of Sound*. San Diego: Academic press, 1992.

A The Affine Motion Model

If $I(n_1, n_2; t)$ is the current image frame and $I(n_1, n_2; t - \tau)$ is the past image, they are related with

$$I(n_1, n_2; t) = I(n_1 - d_1(n_1, n_2), n_2 - d_2(n_1, n_2); t - \tau) = I(n'_1, n'_2; t - \tau) \quad (73)$$

where $d_1(n_1, n_2)$ and $d_2(n_1, n_2)$ are the optical flows in the n_1 -direction and n_2 -direction, respectively. For affine modeled motion, the new pixel co-ordinates (n_1, n_2) are related to the old pixel co-ordinates (n'_1, n'_2) by

$$\begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} n'_1 \\ n'_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \mathbf{A}\mathbf{n}' + \mathbf{b} \quad (74)$$

so that $[d_1(n_1, n_2), d_2(n_1, n_2)]^T = (\mathbf{A} - \mathbf{I})\mathbf{n}' + \mathbf{b}$. Uniform motion (tilt) is a special case of affine motion where $\mathbf{A} = \mathbf{I}$. That is, translation of an object is represented exclusively by \mathbf{b} . Specifically, b_1 represents translation in the n_1 -direction and b_2 represents translation in the n_2 -direction.

The affine motion matrix A has four elements and thereby four degrees of freedom available for describing motion. Therefore, it is expected that A can represent 4 different types of motion. This is indeed true, and these motions are

1. Dilation (expansion or contraction) of the n_1 -direction. The *basic affine matrix* which does (only) this is

$$A_1 = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix} \quad (75)$$

where α is the degree of expansion ($\alpha > 1$) or contraction ($\alpha < 1$). That is, an α of 2 causes the n_1 -direction to expand by 100%, and an α of 1/2 causes the n_1 -direction to shrink by 50%.

2. Dilation (expansion or contraction) of the n_2 -direction. The *basic affine matrix* which does (only) this is

$$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & \beta \end{bmatrix} \quad (76)$$

where β is the degree of expansion ($\beta > 1$) or contraction ($\beta < 1$). β is analogous to α in A_1 .

3. Rotation of the co-ordinate axis by an angle θ in the clockwise direction. The *basic affine matrix* for performing (only) this is

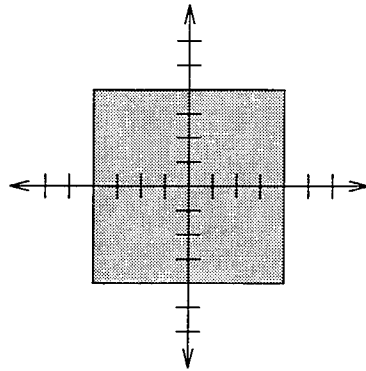
$$A_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \quad (77)$$

4. Skewing of the co-ordinate axis so that the angle between the n_1 -axis and n_2 -axis is the angle ϕ . The two *basic affine matrixes* for doing (only) this are

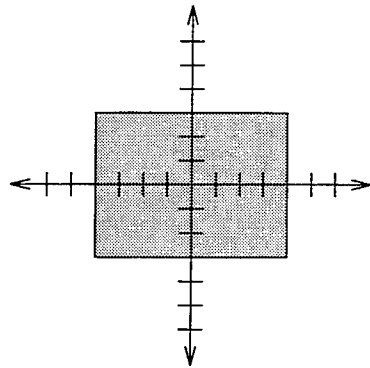
$$A_\phi = \begin{bmatrix} \csc(\phi) & \cot(\phi) \\ 0 & 1 \end{bmatrix}, \quad A_\phi^{-1} = \begin{bmatrix} \sin(\phi) & -\cos(\phi) \\ 0 & 1 \end{bmatrix} \quad (78)$$

Note that the second is the matrix inverse of the first, and therefore the notation A_ϕ^{-1} .

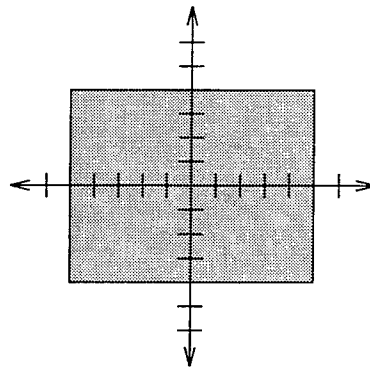
Examples of each of the basic affine motions are illustrated in Figure 16. Each basic matrix is an identity matrix when $\alpha = 1$, $\beta = 1$, $\theta = 0$, or $\phi = \pi/2$, as appropriate. When A is an identity matrix no scaling, rotation, or skewing of the image occurs. Only translation may exist.



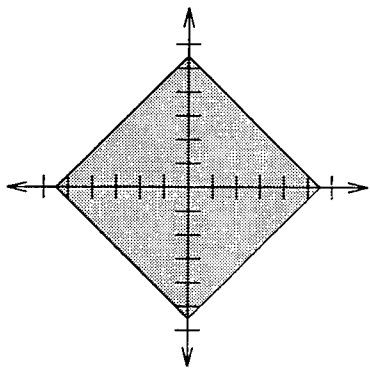
(a) Original 8×8 square image



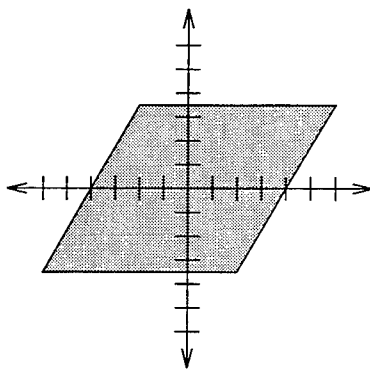
(b) A_1 ($\alpha = 0.75$)



(c) A_2 ($\beta = 1.25$)



(d) A_θ ($\theta = \pi/4$)



(e) A_ϕ ($\phi = \pi/3$)

Figure 16: Examples of affine motion on a 8×8 filled square.

A.1 Compound Affine Motions

The affine motion matrix A does not restrict motion to being only one of the basic affine motions. It allows the motion to be more complicated: an object can both rotate and expand. The A matrix which performs this type of motion can be formed as the product of the basic A matrices A_1 , A_2 , A_θ and A_ϕ . However, the order of the products is important. Rotating an object by 45 degrees clockwise and then expanding its n_1 -axis by a factor of 2 ($\theta = \pi/4$, $\alpha = 2$) is **very different** from expanding the n_1 -axis by a factor of 2 and then rotating 45 degrees.

B N -Point Central Difference

This MATLABTM program finds the N -point central difference gradient estimator weights in vector \mathbf{x} . The program is called with N equal to the number of data points being used in the polynomial fit (N always odd).

```
% N is # points of central difference
```

```
n = (N+1)/2;
```

```
% Generate L matrix
```

```
L = zeros(N);
```

```
for i = 1:N,
```

```
    p = 1;
```

```
    s = 1;
```

```
    for j = 1:N,
```

```
        if i ~= j,
```

```
            p = conv(p,[1 j-n]);
```

```
            s = s * (i-j);
```

```
        end;
```

```
    end;
```

```
    L(i,:) = p/s;
```

```
end;
```

```
% Compute coefficients
```

```
x = L(:,N-1)';
```

C N -point Skewed Difference

This MATLABTM program finds the N -point skewed difference gradient estimator weights in vector \mathbf{x} . The program is called with N equal to the number of data points being used in the polynomial fit.

```
% N = # points, generate L matrix
```

```
L = zeros(N);
```

```
for i = 1:N,
```

```
    p = 1;
```

```
    s = 1;
```

```

for j = 1:N,
    if i ~= j,
        p = conv(p,[1 j-1]);
        s = s * (j-i);
    end;
end;
L(i,:) = p/s;
end;

% Compute coefficients

x = L(:,N-1)';

```

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES
IN SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski
Research Associate
Department of Chemistry

California State University, Northridge
18111 Nordhoff Avenue, CHEM
Northridge, CA 91330

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Phillips Laboratory
Edwards Air Force Base, CA

December 1995

THEMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski
Research Associate
Department of Chemistry
California State University, Northridge

Abstract

In an attempt to develop chemical propellants which would exceed the performance of the current state of the art liquid oxygen/liquid hydrogen system, the concept of using atomic and molecular radicals as high energy additives to cryogenic solid molecular hydrogen (SMH) fuels has been reexamined. This report details efforts to produce and characterize samples of prototypical SMH fuels which will yield visible thermoluminescence (TL) from radical recombination upon warm up. The TL emission spectra and the total TL intensity were recorded during warm up for the simple gases N_2 , O_2 , CO, CO_2 , and CH_4 , as well as various mixtures of N_2 with the others. N_2 was the only gas which yielded TL by itself, while O_2 and CO were the only gases to yield new TL emissions in N_2 mixtures. The transitions evident in the TL emission spectra have been assigned and a kinetic scheme has been proposed. The strongest feature in these spectra was assigned to the $A^3\Sigma_u^+ - X^1\Sigma_g^+$ Vegard-Kaplan bands of molecular nitrogen. Along with an N atom emission at 521.4 nm due to the $2p^3 (^2D) - 2p^3 (^4S)$ transition, a peak at 556.6 nm, assigned to the $2p^4 (^1S) - 2p^4 (^1D)$ transition of the O atom was also observed and attributed to an air impurity. The total TL intensity versus warmup time, or TL glowcurve, was fit to a first-order kinetic model for both the N/D_2 and N/N_2 matrices. The TL glowcurve for N/N_2 matrices contained three peaks which were fit using three sets of N atom detrapping frequency factors and activation energies. The single peak for the N/D_2 TL glowcurve was fit using a single frequency factor and activation energy.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Introduction

This study was carried out to determine which simple gases can be discharged as radicals and co-deposited with D₂ at 3 K in order to produce materials which will yield visible optical emissions upon warm up, i.e., thermoluminescence (TL). These materials constitute prototypical High Energy Density Matter (HEDM) samples and have been used in laboratory-scale experiments to develop the new techniques required for producing, storing, and handling this new class of advanced propellants. As such, this study is part of a greater effort at the Phillips Laboratory at Edwards Air Force Base to adapt HEDM basic research¹ to applied research by coordinating it with the engineering aspects inherent in rocket propellant studies. The interest in this area is from the need for chemical propellants capable of surpassing the performance of the systems currently used. One concept is to use atomic and molecular radicals as high energy additives to cryogenic solid molecular hydrogen (SMH) propellants. These new advanced propellants are expected to improve upon the performance of the current state of the art liquid oxygen/liquid hydrogen (LOX/H₂) system.

The performance of a propellant is characterized by the specific impulse, I_{sp} , which measures the momentum transferred to the vehicle per mass of propellant expelled, and is proportional to the mean velocity of the exhaust². The specific impulse is related to the specific enthalpy of the chemical reactions taking place in the engine, ΔH_{sp} , and to the mean molecular mass of the expelled propellant working fluid, m , by the following equation:

$$I_{sp} \propto \left(\frac{\Delta H_{sp}}{m} \right)^{1/2}$$

Because of this relationship, the lowest molecular mass working fluid is desired as are those additives which increase the ΔH_{sp} . Hence, H₂ is ideal in the capacity as working fluid and matrix host. A survey of combustion energies of atoms and molecules with oxygen was conducted by Carrick³ and indicates which additives will theoretically lead to significant improvements in the specific impulse when doped in SMH.

The gases which were determined to be favorable candidates as precursors for high energy additives were evaluated experimentally in this study using standard matrix-isolation spectroscopy (MIS) techniques. The successful candidates from these MIS studies have been subsequently used in scale-up

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

experiments in which samples are prepared by deposition directly onto a superfluid helium surface.⁴ These subsequent experiments have produced a laboratory quantity of approximately one gram (ten cubic centimeters) of the doped SMH. By taking what is learned from the MIS experiments and expanding upon it, this research helped to determine the relevant physical and chemical properties of the HEDM cryosolid prototypical fuel. Of particular interest is the energy content and thermal properties which are valuable in resolving the issues of how to make, store, and handle the fuel.

The earliest experiments known on the spectra of radicals trapped in the solid state were performed independently by Vegard⁵ and by McLennan and Shrum⁶ in 1924. Their findings caused a considerable interest in this area and became the subject of many studies by these groups. After this initial attention, little was carried out on radicals until the work of the National Bureau of Standards (NBS) Free Radical Program from 1956 to 1959 by Broida and his co-workers.⁷ This work on the emission and absorption spectra of free radicals trapped in solids at low temperatures led to the conclusion that these additives would theoretically improve fuel performance.

The advent of the MIS technique again led to further studies conducted on these radicals at low temperatures with the emphasis on various rare gas matrices.⁸ Hydrogens as matrix hosts have not been frequently studied because of their high vapor pressure at the liquid helium boiling temperature.⁹ But, this deterrent can be overcome by pumping the liquid helium cryogen which lowers the temperature sufficiently to work in SMH matrices. Another problem of working with SMH, not incurred with the rare gas matrices, is the possibility of the SMH host reacting with the guest radicals. By insuring the reactions of the energetic species with molecular hydrogens are endothermic or have at least a slight activation barrier, this is no longer an obstacle. To date, many atoms including Al, B, N, and O have been successfully trapped in SMH matrices¹⁰⁻¹².

Although more information is being amassed on this subject all the time, there are still many questions which remain unanswered. Among the questions we hope to address in this work include: 1) which radical species can be trapped in D₂?; 2) which species will react with the matrix host; 3) which species will yield TL; and 4) what is the mechanism of TL?

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Experimental

The experimental apparatus for the MIS experiments is depicted in Figures 1 and 2. Figure 1 is a schematic diagram of the vacuum and gas handling system employed in the experiments. Figure 2 represents the data acquisition and sample area of the cryostat.

A liquid helium transfer cryostat (RG Hansen High Tran) with a mechanical pump (Leybold Model D40B) was used to lower the cold tip temperature to about 3 K by pumping on the liquid helium exit port. Temperature readings were made with a temperature controller (Scientific Instruments Model 5500-1-25) connected to two silicon diode temperature sensors inside the cryostat. One of the sensors was mounted directly on the cold tip and the other on the far end of the sample substrate holder. An average was taken from these two readings for the reported matrix temperature.

Cryostat pressures of 10^{-8} torr were obtained prior to cool down using a turbo molecular pump (Leybold NT 150/360) backed by a mechanical pump (Varian SD-200) (see Figure 1). A cold cathode ionization gauge (Varian 860A) and a Residual Gas Analyzer (RGA) (Spectra Vacscan) were used to determine the total pressure in the lines and the identity of the gases present in the vacuum chamber. The RGA showed the main residual gases to have charge-to-mass ratios of 28, 18, 32, 17, 44, and 1 and were attributed to air (N_2 , O_2 , CO_2) and water (H_2O , OH, and H) impurities, respectively.

The experiments were conducted using technical and research grade gases (Air Products for D_2 (99.0% purity) and CH_4 (99.0% purity); Matheson for CO (99.5% purity); and Alphagaz for N_2 (99.995% purity), O_2 (99.999% purity), and CO_2 (99.995% purity). Mixtures of gases were made in the vacuum lines by alternately adding the gases and letting stand between 30 to 90 minutes to encourage mixing. The oxygen impurity occurred from an air leak in the vacuum line above the microwave discharge.

The gases were deposited onto a cold sapphire plate fixed at a 45° angle from the base (see Figure 2). The guest matrix gas flowed through a microwave discharge (OPTHOS) set at a power level between 40 and 100 W (OPTHOS Power Supply). The microwave discharge pressure was kept between 100 and 150 mtorr. The microwave discharge consisted of a 12 mm diameter quartz tube surrounding a 6 mm diameter quartz tube with a 0.1 mm diameter pinhole opening. The outer tube was cooled by a force flow of N_2 gas during the discharge. A Tesla coil

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

was used to "ignite" the microwave discharge. The discharge formed radicals from the sample gas which flowed through the pinhole opening and deposited onto the matrix window. The host D_2 gas was deposited perpendicular to the window.

The matrix consisted of about 5% of the sample gas in the D_2 matrix. The composition was controlled by the gases' flow rates which were established by filling the two separate vacuum lines with the guest and host gases. The composition was controlled during the deposition by adjusting the needle valves using the pressure readings from the two manifolds (PDR-C-2C Power Supply Readout for D_2 and Datametrics 1500 Power Supply Readout for the microwave flow rate). The flow rates were maintained at about 0.1 mmol/hr for the guest and 2.0 mmol/hr for D_2 gas.

After a 30 minute deposition, the microwave discharge power was shut off and the gas flow was stopped by closing the needle valves. The matrix was warmed up by shutting off the liquid helium flow or by heating with the temperature controller. The latter allowed for a slower and steadier warm up of the matrix, but only gave a maximum temperature of 25 K. The temperature was recorded from the analog output of the temperature controller by the source compensation channel of the Optical Multichannel Analyzer (OMA) (EG&G Model 1460) and an analog input on the lock-in amplifier with a buffer preamp (Stanford Research Systems SRS Analog Processor SR235) to increase the temperature voltage signal ten-fold. The OMA monochromator (EG&G Model 1234) with a low resolution grating (150 grooves/mm, $\Delta\lambda=1.5$ nm full width at half max) and a detector with a wavelength range of 230 to 830 nm (EG&G IR 1421 Array Detector) were used to record the TL emission spectrum at 1 sec/scan.

The TL emission was focused point-to-point by a 2 inch focal lens placed 4 inches from the detector and the matrix (see Figure 2). The light passed through a mechanical rotating chopping wheel attached to a controller set at 408 Hz (Stanford Research Systems SR540 Chopper Controller). The light continued through a 80 μ m slit to a photomultiplier tube (PMT) (EMI 1P28A) using a bias voltage of -600 V (Stanford Research Systems Model PS 325 High Voltage Power Supply). The PMT converts light into an electrical signal. The slit width and bias voltage value were chosen so the strongest TL would not damage the PMT. A bipolar preamp (Stanford Research Model SR552) increased the signal ten-fold before going to a lock-in amplifier (Stanford Research Systems SR530). An oscilloscope (Tektronix 2440) visually monitored the intensity of

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

the signal. The chopper and lock-in amplifier combination improved the signal-to-noise ratio relative to a simple DC electrometer setup.

This signal was finally transferred from the lock-in amplifier to a PC via a GPIB interface. A BASIC program was used to record the intensity of the in phase signal of the TL glowcurve, the out phase signal, the time, and the temperature.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

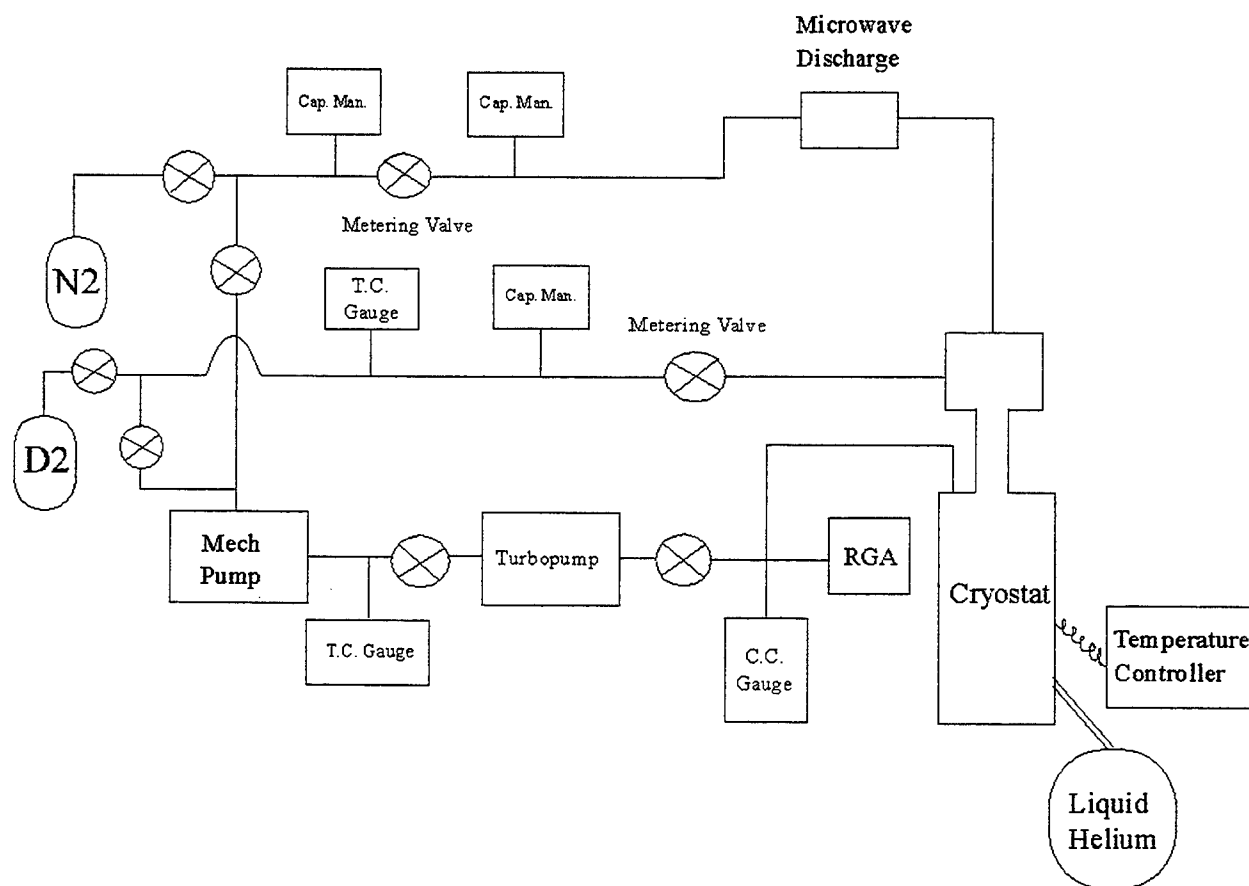


Figure 1. Schematic diagram of the vacuum and gas handling system used in the MIS experiments to trap radicals in SMH matrices. ⊗ represents the valves in the vacuum line, RGA is the Residual Gas Analyzer, C.C. Gauge stands for the cold cathode ionization gauge, T.C. Gauge is the thermocouple gauges, Mech Pump represents the mechanical pump, and Cap. Man. is the capacitance manometers. See text for further explanation.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

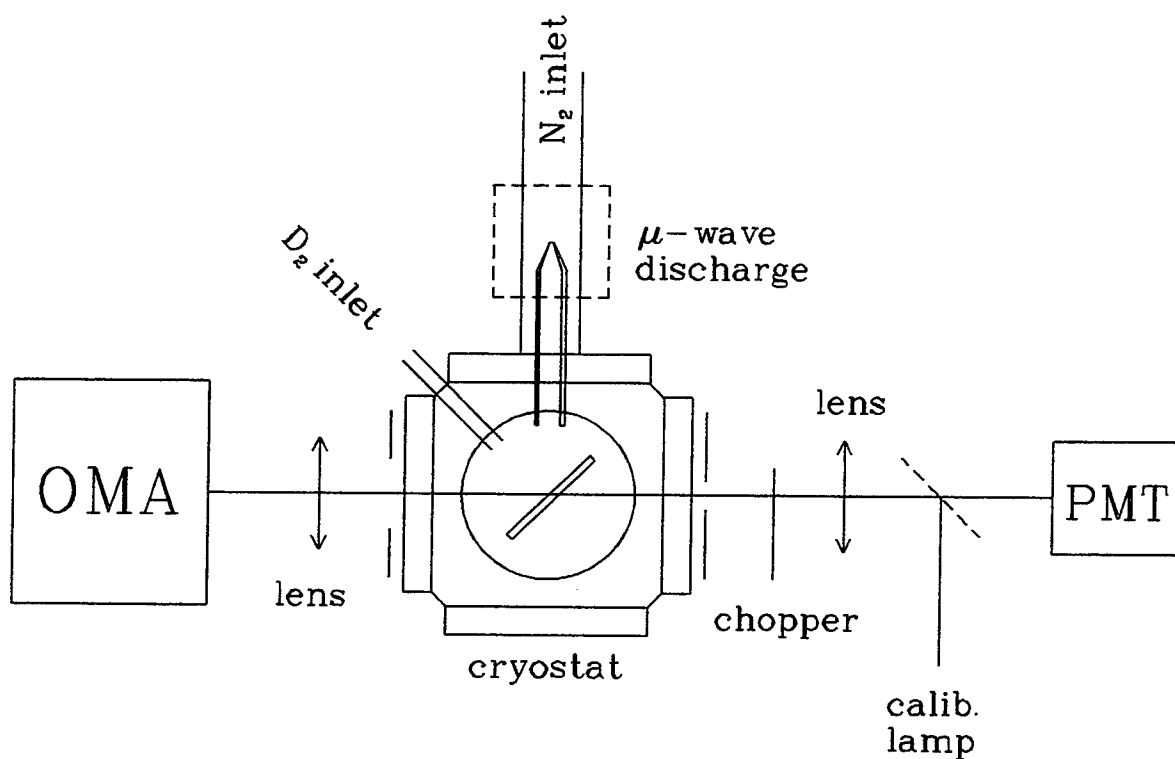


Figure 2. Schematic representation of the cryostat, data acquisition system, and sample region from Figure 1. OMA represents the Optical Multichannel Analyzer, the PMT is the photomultiplier tube, and the calib. lamp depicts the area where the mercury calibration lamp was placed. See text for further explanation.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Results

The gases which were discharged and trapped in matrices were N_2 , O_2 , CO , CH_4 , CO_2 and mixtures of N_2 in varying dilutions with each of the other gases. Intense TL was detected only from the nitrogen-containing matrices. It was not observed from the other gases unless nitrogen was present. All the transitions evident in the TL emission spectrum have been previously assigned and are listed in Table 1. These experiments were performed in solid nitrogen matrices and in SMH matrices to compare with other experiments performed previously by Fajardo et al. in their work at Phillips Laboratory¹¹. The observed intensity of the TL was much less in the D_2 matrix than the N_2 matrix.

Low temperature phosphorescence was observed as a visible green glow from some samples immediately following the end of the deposition period and before the sample was warmed up to the maximum temperature. In the N/N_2 matrix experiments, the intensity of the green phosphorescence decayed exponentially with a 35 (± 5) second lifetime; the emission from the N/D_2 matrices was much weaker and decayed with a 30 (± 5) second lifetime. This green light was observed previously and attributed to³ the $N(^2D) - (^4S)$ atom emission at 521.4 nm which had a lifetime of 29 seconds in solid nitrogen.¹³ Thermal explosions were observed in some experiments as flashes of green light emitted during the deposition and warming up process. These were especially prevalent in the N/N_2 matrix experiments.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Section A: Solid Nitrogen Matrices

Figure 3 depicts TL from N/N₂ matrices. Figure 3a exhibits TL from a N/N₂ matrix deposition. It shows two N₂ molecular transitions in this emission spectrum; the first is the more intense of the two is known as the Vegard-Kaplan band⁵ and is the $A^3\Sigma_u^+ \rightarrow X^1\Sigma_g^+$ transition which occurs in the range of 300 nm to 460 nm. The second weaker set from 600 nm to 760 nm is due to the $B^3\Pi_g \rightarrow A^3\Sigma_u^+$ transition. The most intense peak at 521.4 nm is the N atom α emission and has been interpreted by Bass and Broida⁷ as the matrix induced, spin and dipole forbidden $2p^3 (^2D) \rightarrow 2p^3 (^4S)$ transition. Figure 3b shows TL from a discharged N/N₂ sample which most likely had an air impurity of an unspecified composition. A feature previously assigned to the O atom appears at 556.6 nm and was interpreted by Broida and Peyron¹⁴ as the matrix induced, dipole forbidden $2p^4 (^1S) \rightarrow 2p^4 (^1D)$ transition, called the β emission. Figure 3c depicts TL from another N/N₂ matrix deposition with a much larger air impurity. Here the N₂ transition bands are absent. Evident between 350 nm and 480 nm is the O₂ $A'^3\Delta_u \rightarrow X^3\Sigma_g^-$ molecular transition.¹² This is a very weak transition since it was only seen when the N and O atom intensities were saturated. The peak at 593.0 is the α' N atom emission and is thought to involve a process in which an electronic transition of a nitrogen atom is accompanied by a simultaneous vibrational transition of a nearby nitrogen molecule.¹⁵ The N atom peak at 793.0 nm is tentatively assigned to the γ transition.¹⁶ A summary of the observed transitions and their assignments is given in Table 1.

Figure 4 shows the dependence of the TL intensity for an N/N₂ matrix upon the time of warming up and upon the temperature. Figure 4a depicts the TL intensity and the associated temperature ramp for the N/N₂ experiment of Figure 3a. The temperature associated with the peak of maximum intensity is 17 K. Figure 4b shows the histogram computed from figure 4a using 2 K wide bins. The emission occurred from 7 K to 23 K with the highest peak at 19 K and minor peaks at 8 K and 12 K. There were other peaks observed in other experiments but these were not always reproducible and were attributed to thermal explosions during which the measured substrate temperature does not track the local sample temperature.

Figure 5 is another depiction of the TL glowcurve of the N/N₂ matrix of Figure 3a obtained from the OMA data by integrating the various emission

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

peaks. This figure shows the relative contribution and lifetimes of each individual species in the emission spectra to the overall TL intensity. Figure 5a is the curve of the O atom intensity contributing about 3.1% to the overall intensity. Figure 5b depicts the N atom intensity of 18.5%. Figure 5c shows the N₂ molecule which is the major contributor at 78.4% of the total. Figure 5d is the total intensity calculated by the summation of the intensities from O atom, N atom and N₂ molecule emissions.

Figure 6 is a variation of the TL glowcurves of Figure 5. The intensities of the three species have been normalized to the same height so as to compare the shape of the curves and to obtain the lifetime of the species. Figure 6a is due to the O atom and Figure 6b is due to the N₂ molecule, both of which demonstrate a similar shape. Figure 6c is the N atom emission which has a longer lifetime than the other species and, hence, a greater relative contribution in the intensity of the TL glowcurve past 40 seconds.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

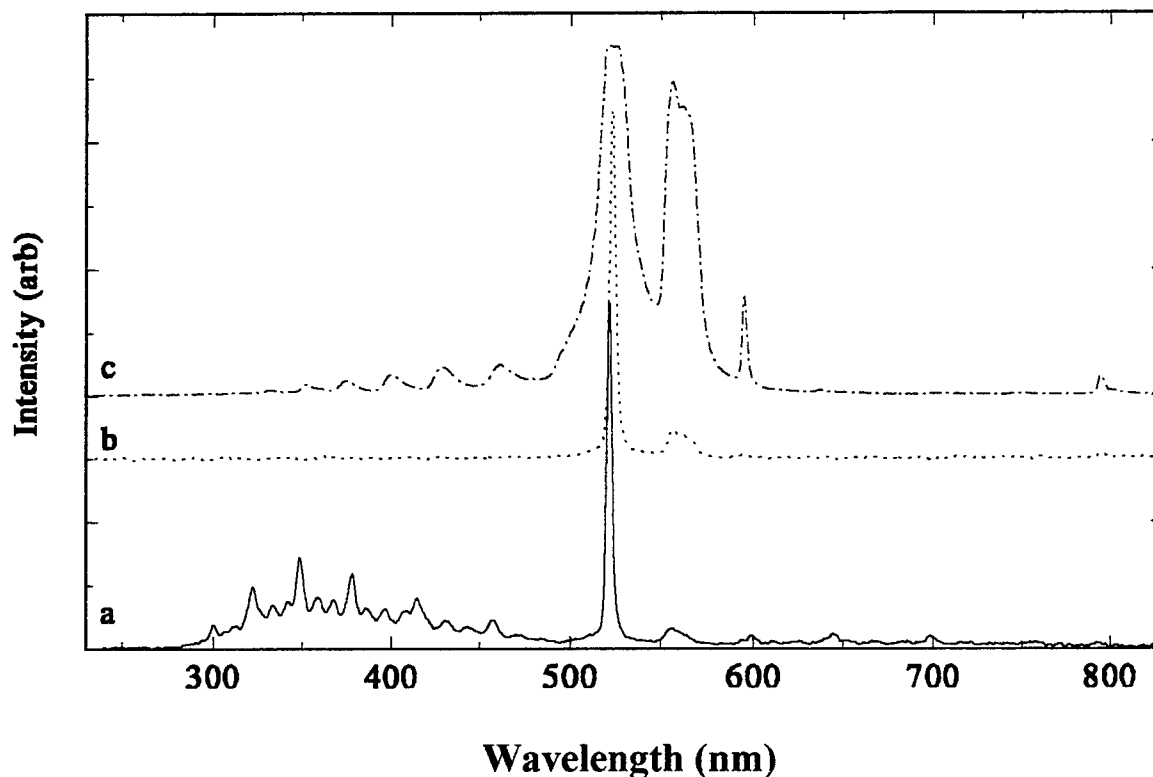


Figure 3. Thermoluminescence from N/N₂ matrix. Trace (a) is from a N/N₂ sample produced with: T = 3.8 K, P_{μwave} = 125 mtorr N₂, discharge power = 50 W, and a 0.09 mmol/hr N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 41 K. Trace (b) is from a N/N₂ sample with an air impurity produced with: T = 4.7 K, P_{μwave} = 400 mtorr N₂, discharge power = 50 W, and a 0.029 mmol/hr N₂ flow rate for the 20 minute deposition; the emission was recorded during warm up from 4 to 41 K. Trace (c) is from a N/N₂ sample with a much larger air impurity produced with: T = 4.7 K, P_{μwave} = 800 mtorr N₂, discharge power = 50 W, and a 0.57 mmol/hr N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 41 K. The spectra were recorded with an optical multichannel analyzer equipped with an intensified array detector. They have not been corrected for the wavelength dependent detector response, and have been arbitrarily normalized for ease of presentation.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
 SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

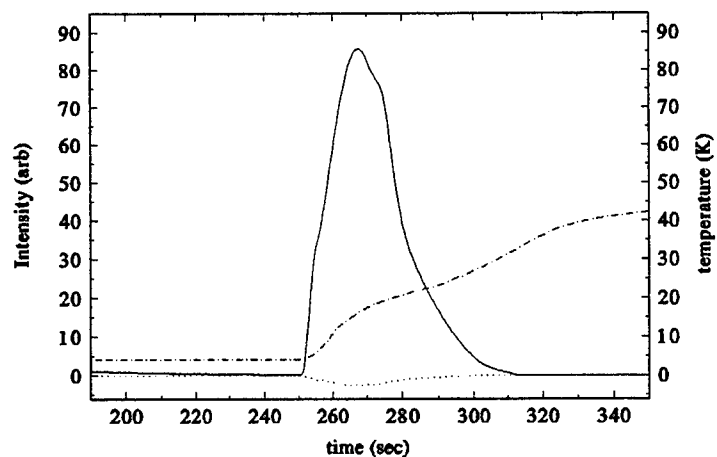


Figure 4a. Thermoluminescence glowcurve from N/N₂ matrix of Figure 3a. (--) is the total TL intensity (in phase signal), (...) is the out phase signal, and (- - -) is the temperature ramp.

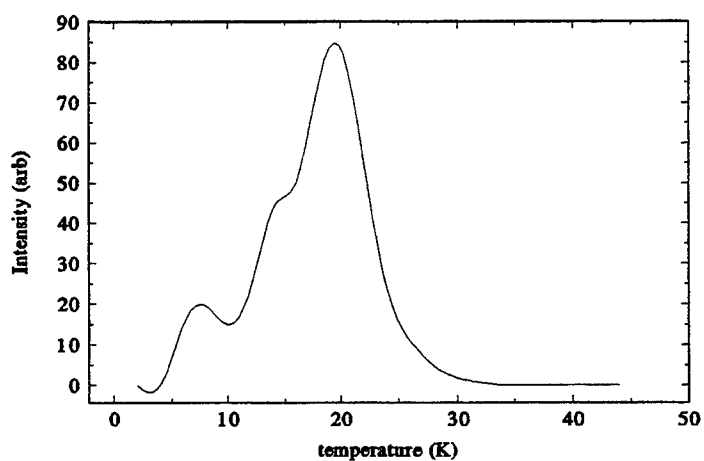


Figure 4b. Histogram of thermoluminescence glowcurve of N/N₂ matrix of Figure 4a.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

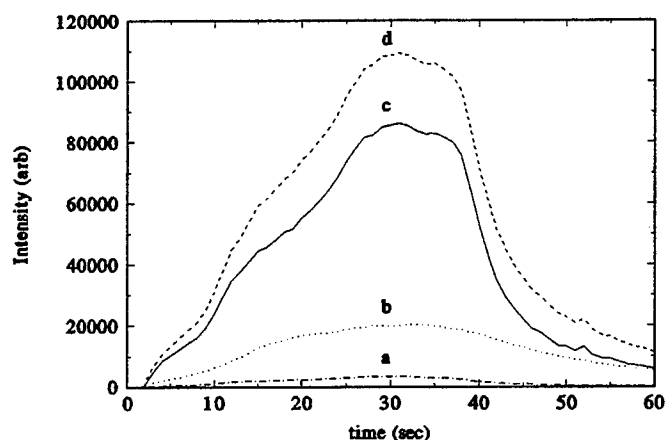


Figure 5. Thermoluminescence glowcurve of the N/N₂ matrix of Figure 3a compared to the total intensity of the emission. Trace (a) is the O atom intensity, trace (b) is the N atom intensity, while trace (c) is the N₂ molecule intensity, and trace (d) is the cumulative total of the three species.

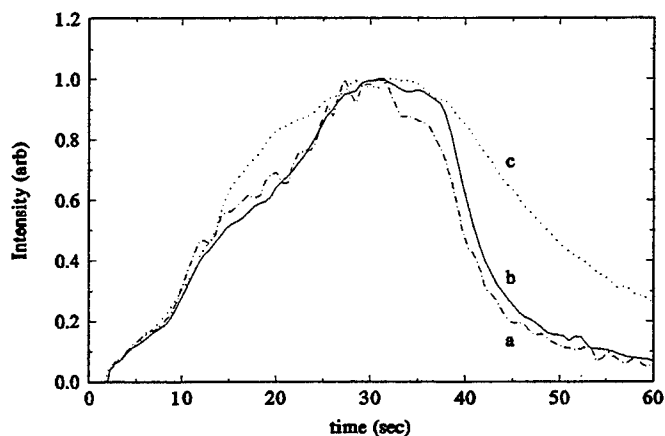


Figure 6. Normalized thermoluminescence glowcurve intensities for the three species in Figure 5. Trace (a) is the O atom intensity, trace (b) is the N₂ molecule, and trace (c) is the N atom normalized curve.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Section B: D₂ Matrices

Figure 7 depicts TL from N/D₂ matrices. Figure 7a shows TL from a N/D₂ matrix deposition. The two N₂ molecular transitions in this emission spectrum match those observed in the previous section; but the N (²D) - (⁴S) atom transition is nearly absent. Figure 7b shows TL from a 10% O₂/N₂ sample co-deposited in a D₂ matrix. The N atom transition for this sample is the same as assigned by Broida and Peyron above. The O (¹S) - (¹D) emission is clearly visible. Figure 7c depicts TL from a N/D₂ matrix deposition which had an air impurity of an unspecified composition. The O atom transition is the strongest band evident and N₂ molecular bands are absent. Both the N atom transition bands are observed as well as emissions not previously seen for the N atom in the other spectra. One of these additional peaks is found at 471.5 nm and is assigned to the α" emission¹³ of the 2p³ (²D) - 2p³ (⁴S) N atom transition. A peak at 359.4 nm is assigned to the μ emission of the 2p³ (²D) - 2p³ (⁴S) N atom transition.¹³ The N atom peak at 793.0 nm is tentatively assigned to the γ transition.¹⁶ This sample most likely contained solid N₂ regions (see TL data in Figure 9c). Figure 7d also depicts TL from a N/D₂ matrix deposition with an air impurity of unspecified composition. Here the N atom transition band is minimal and the N₂ transition bands are absent. The O atom transition is the most dominant band. A summary of all the observed transitions and their assignments is given in Table 1.

Figure 8 shows the dependence of the TL intensity upon the time of warming up for the N/D₂ matrix. Figure 8a depicts the TL intensity and the associated temperature ramp for the N/D₂ experiment of Figure 7a. The temperatures of the three peaks are 5 K, 7 K, and 8 K. Figure 8b is a histogram computed from the data in Figure 4a using 1 K wide bins; the intensities were then plotted against the temperature. The primary emission occurs from 4 K to 9 K with the major peak being at 7 K and minor peaks at 11 K and 13 K. There were also other emissions in some experiments which were not always reproducible in the glowcurves and were attributed to thermal explosions or to emissions from segregated solid nitrogen regions.

Figure 9 is the histograms of all the experiments of Figure 7 and demonstrates the reproducibility of the experiments. All of the histograms depicted have a peak at approximately 7 K. The experiments showing evidence of N atom transition had an extra peak at approximately 5 K. The histograms

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

for Figure 9 were calculated the same as for Figure 8b. Figure 9a is the same histogram as in Figure 8b. Figure 9b shows the histogram computed from the TL glowcurve data from Figure 7b. The primary emission occurs from 4 K to 11 K with the major peaks being at 4 K and 7 K. Figure 9c shows the histogram computed from the TL glowcurve data from Figure 7c. The primary emission occurs from 4 K to 9 K with the most intense peak being at 5 K and minor peaks at 7 K, 12 K, and 16 K. Figure 9d shows the histogram computed from the TL glowcurve data from Figure 7d. The primary emission occurs from 4 K to 9 K with the major peak being at 7 K and a minor peak at 4 K.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

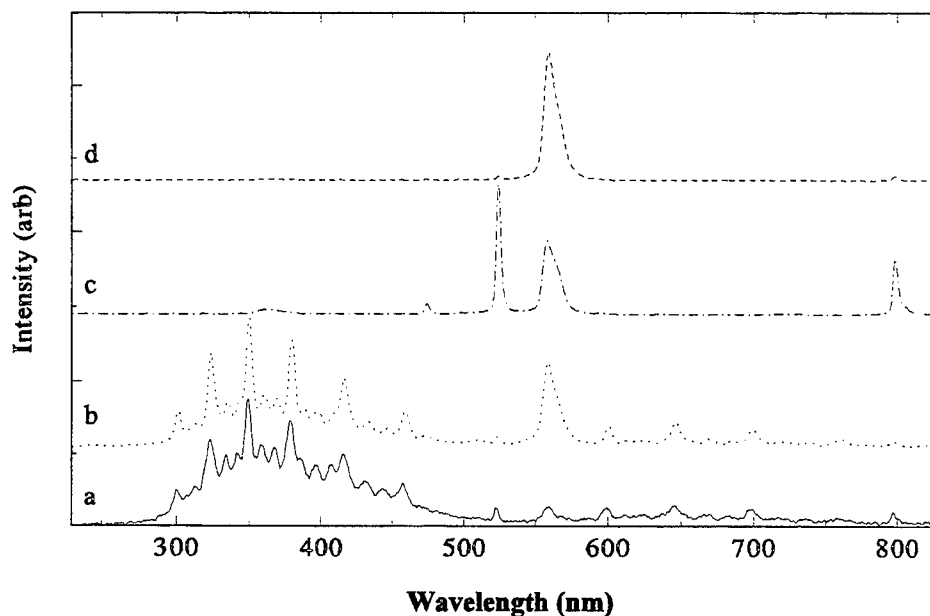


Figure 7. Thermoluminescence from N/D₂ matrices. Trace (a) is from a N₂/D₂ sample produced with: T = 3.6 K, P_{μwave} = 125 mtorr N₂, discharge power = 60 W, a 1.9 mmol/hr D₂ flow rate, and a 0.10 mmol/hr N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 24 K. Trace (b) is from a 10% O₂/N₂/D₂ sample produced with: T = 3.6 K, P_{μwave} = 125 mtorr O₂/N₂, discharge power = 60 W, a 1.8 mmol/hr D₂ flow rate, and a 0.10 mmol/hr O₂/N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 24 K. Trace (c) is from a N/D₂ sample with an air impurity produced with: T = 3.5 K, P_{μwave} = 350 mtorr N₂, discharge power = 40 W, a 2.0 mmol/hr D₂ flow rate, and a 0.27 mmol/hr N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 24 K. Trace (d) is from a N/D₂ sample with an air impurity produced with: T = 3.7 K, P_{μwave} = 100 mtorr N₂, discharge power = 40 W, a 1.9 mmol/hr D₂ flow rate, and a 0.03 mmol/hr N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 24 K. The spectra were recorded with an optical multichannel analyzer equipped with an intensified array detector. They have not been corrected for the wavelength dependent detector response, and have been arbitrarily normalized for ease of presentation.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

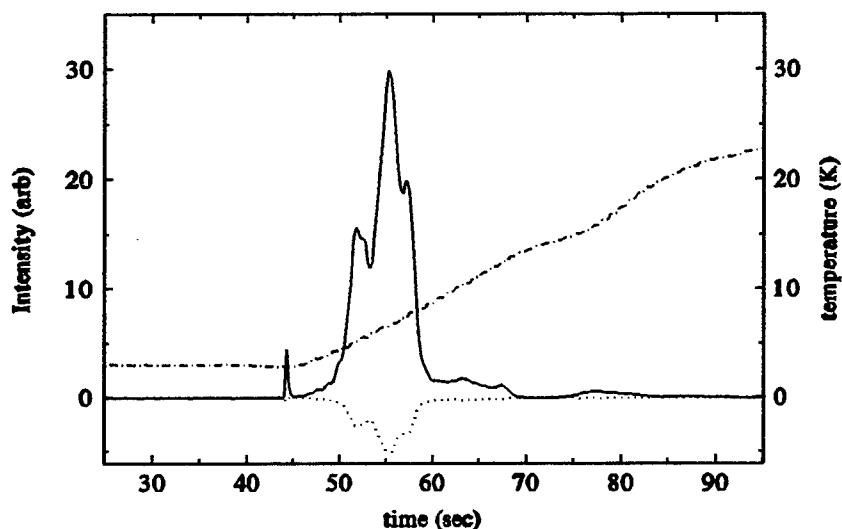


Figure 8a. Thermoluminescence glowcurve from N/D_2 matrix of Figure 7a. (---) is the total TL intensity (in phase signal), (...) is the out phase signal, and (- - -) is the temperature ramp.

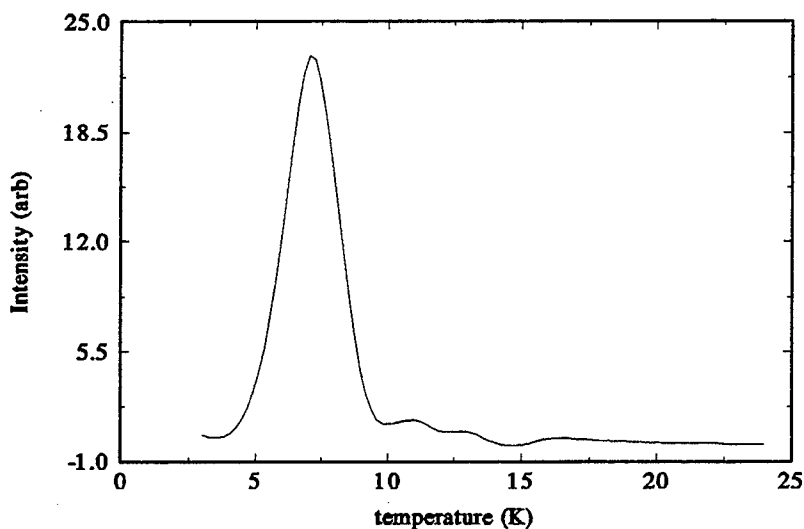


Figure 8b. Histogram of thermoluminescence glowcurve of N/D_2 matrix of Figure 7a.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

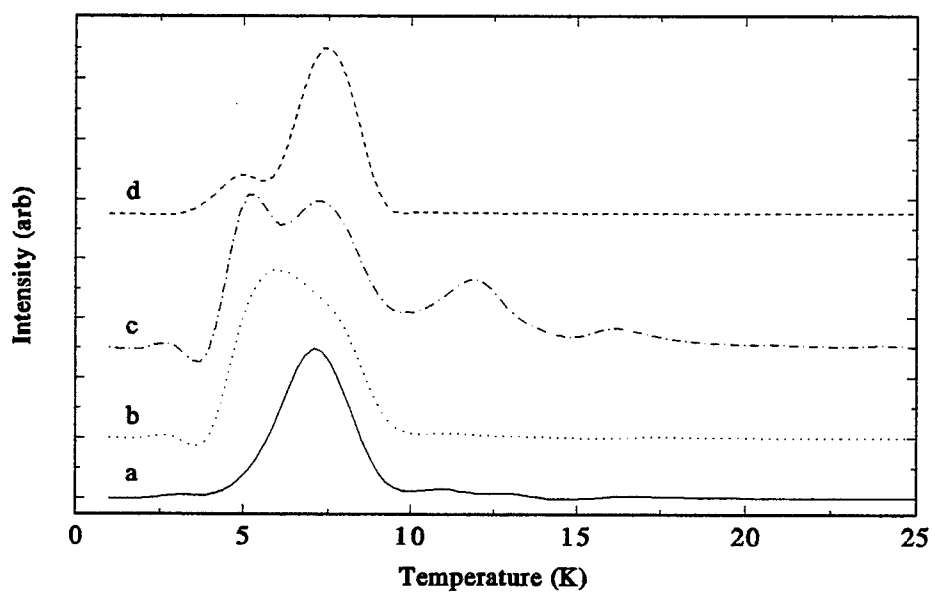


Figure 9. Histograms of thermoluminescence glowcurves for N/D_2 matrices of Figure 7. Trace (a) is from Figure 7a. Trace (b) is from Figure 7b. Trace (c) is from Figure 7c. Trace (d) is from Figure 7d. They have been arbitrarily normalized for ease of presentation.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Section C: Mixtures of Nitrogen with Other Gases in D₂ Matrices

In other gases in D₂ matrices, no other species were trapped that thermoluminesced or chemiluminesced or accepted energy transfer from N₂⁺ or N⁺ individually in the matrix. When O₂ was mixed with varying amounts of N₂ (at least 50% N₂ was needed), O atom β emission at 556.6 nm, due to the matrix induced, dipole forbidden $2p^4 (^1S) \rightarrow 2p^4 (^1D)$ transition,¹² was observed in addition to the Vegard-Kaplan band⁵ of the N₂ molecule, the $A^3\Sigma_u^+ \rightarrow X^1\Sigma_g^+$ transition which occurs in the range of 300 nm to 460 nm. However, N atom emission was not detected (see Figure 7b). CO/N₂ mixtures (50:50 ratio) also yielded TL and the same emission spectra as seen for O₂/N₂/D₂. Again, no N atom emission was detected, nor was the carbon emission (the C₂ Swann bands) observed. CH₄ and CO₂ did not give any emission, though CH₄ matrix did show the Vegard-Kaplan band transition of the N₂ molecule with a 90% N₂/CH₄ mixture only. In the cases where TL emission was detected with the mixtures, the intensity of the bands was greatly diminished as compared to the N/D₂ matrices. Figure 10 depicts the emission spectra of a 50:50 CO/N₂ mixture and a 90:10 N₂/CH₄ mixture. All of the bands evident in the spectra are listed in Table 1.

Discoloration of the microwave discharge chamber was observed when CH₄ or CO experiments were carried out. The chamber turned black by the end of a 30 minute deposition of CO, a process which was attributed to a carbon build-up on the quartz tube. The CH₄ experiments yielded a brown layer primarily on the top section of the chamber tube. This layer was thought to be a hydrocarbon polymer. Also, during the CH₄ runs the microwave power level setting was extremely difficult to maintain for the entire deposition time and had to be watched carefully to prevent the microwave discharge from extinguishing. It is possible that the N⁺ or N₂⁺ species were quenched by reactions with the microwave discharge cavity walls during these experiments with other gases (see Discussion, Section A).

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

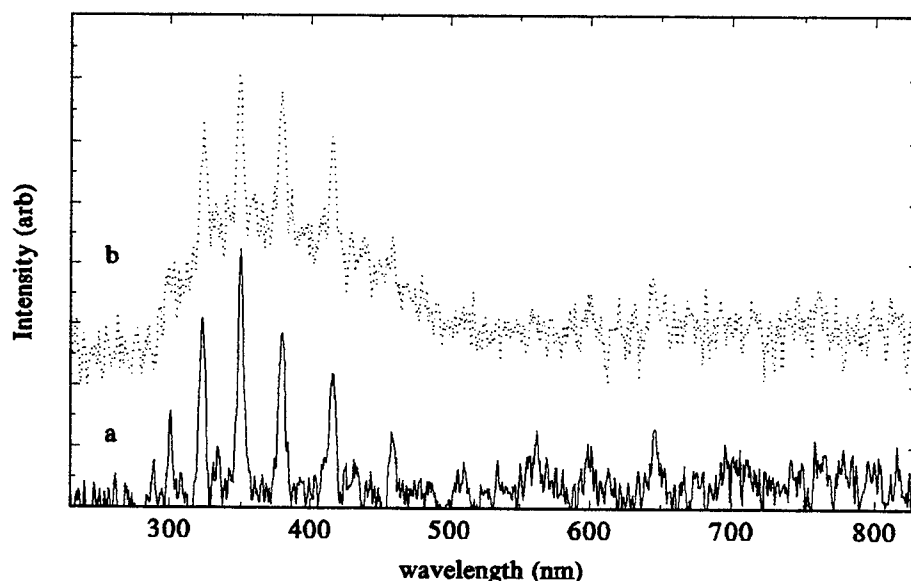


Figure 10. Thermoluminescence from separate mixtures of CO and CH₄ with N/D₂ matrices. Trace (a) is from a 50:50 ratio of CO/N₂ in D₂ sample produced with: T = 3.7 K, P_{uwave} = 125 mtorr CO/N₂, discharge power = 90 W, a 1.7 mmol/hr D₂ flow rate, and a 0.10 mmol/hr CO/N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 24 K. Trace (b) is from a 10:90 ratio of CH₄/N₂ in D₂ sample produced with: T = 3.7 K, P_{uwave} = 125 mtorr N₂, discharge power = 60 W, a 2.2 mmol/hr D₂ flow rate, and a 0.23 mmol/hr N₂ flow rate for the 30 minute deposition; the emission was recorded during warm up from 4 to 24 K. The spectra were recorded with an optical multichannel analyzer equipped with an intensified array detector. They have not been corrected for the wavelength dependent detector response, and have been arbitrarily normalized for ease of presentation.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Table 1. Summary of spectra emitted in matrices of N/N₂ and N/D₂.

Emitting Species	NBS Designation	Assignment	λ (nm)	Literature λ (nm) ^{5,7,12-16}
N	α	$^2D \rightarrow ^4S (0,0)$	521.4	522.9 ⁷⁾
	α'	$^2D \rightarrow ^4S (0,1)$	593.0	594.5 ¹⁵⁾
	α''	$^2D \rightarrow ^4S (1,0)$	471.5	469.9 ¹³⁾
	μ	$^2P \rightarrow ^4S (0,0)$	359.4	347.9 ¹³⁾
	?	$^2D \rightarrow ^4S$	793.0	793.0 ¹⁶⁾
N ₂	VK	$A^3\Sigma_u^+ \rightarrow X^1\Sigma_g$	300 - 460	230 - 445 ⁵⁾
	? G	$B^3\Pi_g \rightarrow A^3\Sigma_u^+$	600 - 760	630 - 677 ⁵⁾
O	β	$^1S \rightarrow ^1D(0,0)$	556.6	554.9 ¹⁴⁾
O ₂	A	$A'^3\Delta_u \rightarrow X^3\Sigma_g^-$	350 - 480	357 - 639 ¹²⁾

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

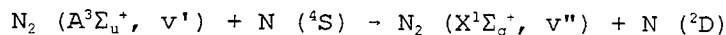
DISCUSSION

Section A: Spectroscopic Assignments

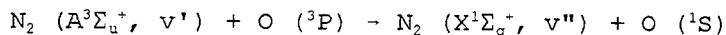
The spectra observed in these experiments of the N/N₂ and N/D₂ matrices reproduced and extended previous work¹¹ through the assignment of the spectral features and the additional gases attempted. The spectroscopic assignments made also agree with this previous work and the prior transition assignments^{5,7,12-16} (see Table 1). The major emitters evident in the spectra are the excited states of N*, N₂*, and O* and O₂*. From these data, it is apparent that N atoms are trapped in the N₂ and D₂ matrices and are a prerequisite before any TL occurs.

The low temperature phosphorescence observed at the end of the matrix deposition process has been attributed to emission by surviving excited N* atoms produced in the discharge.⁷ After a few minutes all low temperature emission from these matrices cease. At this point, it is highly unlikely that any other electronically excited species exist trapped in the matrix. Thus all the subsequent observed TL emitting species must derive their electronic excitation from recombination of ground state N (and/or O) atoms. For example, the excited N₂* (A³Σ_u⁺) species can be formed by recombination of two N (⁴S) atoms, and the O₂* (A³Δ_u) state can be populated during the recombination of two O (³P) atoms.¹⁷ However, the emissions observed from the N₂* (B³Π_g) state does not correlate with two N (⁴S) atoms, and from excited N* (²D) and O* (¹S) atoms, implies that more complicated processes are at work in the formation of some excited state species.

One complex process that has been proposed to account for excited state formation is energy transfer from an excited N₂* (A³Σ_u⁺) molecule to a ground state N (⁴S) atom, a well known process in the gas phase:¹⁵



For v'=0, this reaction is exothermic up to v''=14. The excited O (¹S) atoms are created by the energy transfer from N₂ (A³Σ_u⁺, v') molecules in analogy to the above process:



Thermoluminescence Studies of Simple Species in
Solid Molecular Hydrogen Matrices

Janet Petroski

With $v'=0$, this reaction is expected to be exothermic up to $v''=6$ or 7 , depending on the precise excitation energy of O ($^1S - ^3P$) in the N_2 matrix.

The efficiency of these energy transfer processes can be amazingly high; Figure 3b shows a TL spectrum from a N/N_2 matrix in which only N^* (2D) and O^* (1S) emissions are apparent. The funneling of all this energy from N_2^* excitations into the N and O atom traps suggests that they are highly mobile and should be thought of as "excitons" which can move easily through the nitrogen solid.¹⁸ The absence of such a mechanism for N_2^* mobility in D_2 matrices may explain why the N^* atom emissions are so much weaker than in N_2 matrices.

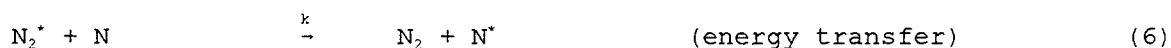
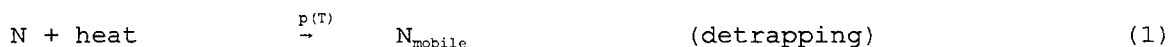
Almost all the spectral emissions evident are from spin and/or electric dipole forbidden transitions. They occur because of perturbations by the matrix which include phonon and vibron coupling. These matrix perturbations modify the initial and final states, breaking their original symmetries and hence lead to violation of the gas phase selection rules. With this model it is possible to characterize the observed emissions as simple combinations of electronic transitions of the nitrogen (or oxygen) atom with the transitions of the N_2 solid. The N atom α emission and the O atom β emission involve coupling of the electronic states with the low frequency lattice phonons of the N_2 solid. The N atom α' emission also involves the creation of a vibrational excitation (vibron) in the N_2 solid, whereas the N atom α'' emission requires the destruction of a vibron.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Section B: TL Kinetics

The kinetics of the TL process were analyzed by assuming a homogeneous matrix of uniform temperature and concentration.¹⁹ This model presented below assumes that the ground state N atoms trapped in the matrix are thermally activated or detrapped to a mobile state (equations 1 and 2). From this state they can recombine with another mobile N atom or another N atom trapped in the matrix to form an excited N_2^* molecule (equations 4 and 5). These mobile N atoms can also be retrapped into the matrix before they can recombine (equation 3). The recombined N_2^* molecule is in an excited state which can then luminesce (equation 6). In the case where the N atom TL is observed, there are the additional steps for the transfer of energy from the N_2^* molecule to an N atom which leads to the N^* atom emission (equations 7 and 8). A kinetic sequence representing these various processes is as follows:



in which $p(T)$ is the temperature dependent detrapping rate constant; A is the rate constant for the retrapping of the N atom, A_r is the rate constant for the recombination of the mobile N atom with one from the matrix; and A_r' is the rate constant for the recombination of two mobile N atoms. The rate constant for the emission of N_2^* is given by $1/\tau_{\text{rad}}$ and for N^* atom by $1/\tau_{\text{rad}}'$. The excited state energy transfer rate constant from a N_2^* molecule to a matrix N atom is expressed as k . Even though the total contribution to the TL

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

intensity due to the N atom was much less than that due to the N₂ molecule, the N atom emission did have a longer lifetime than the N₂ in the N/N₂ matrix, thereby donating a greater relative contribution to the intensity of the TL glowcurve past 40 seconds (see Figures 5 and 6).

The following system of first-order, non-linear, coupled differential equations were derived from the above mechanism:

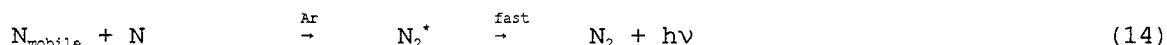
$$dN/dt = -p(T)N + AN_{mobile} - A_rNN_{mobile} - kN_2^*N + 1/\tau_{rad} \cdot N^* \quad (9)$$

$$dN_{mobile}/dt = p(T)N - AN_{mobile} - A_rNN_{mobile} - \frac{1}{2}A_rN_{mobile}^2 \quad (10)$$

$$dN_2^*/dt = A_rNN_{mobile} + \frac{1}{2}A_rN_{mobile}^2 - 1/\tau_{rad}N_2^* - kN_2^*N \quad (11)$$

$$dN^*/dt = kN_2^*N - 1/\tau_{rad}N^* \quad (12)$$

Given values for all the microscopic rate constants, a numerical solution of this system of equations would be straightforward. However, no direct independent information about any of these parameters is available, except for the emission radiative lifetimes. Furthermore, various moderate approximation techniques were applied in an unsuccessful effort to solve these equations. A standard approach was finally used, as was done in previous TL studies, in which the assumption was made that the initial detrapping step is rate determining so that most of the subsequent kinetics can be ignored. Ignoring the kinetics of the retrapping (3), energy transfer (7), and radiative (6 and 8) steps, and imposing a steady state condition on the mobile N atoms, a simplified kinetic scheme is obtained:



The corresponding rate equations are as follows:

$$dN/dt = -p(T)N - A_rNN_{mobile} \quad (15)$$

$$dN_{mobile}/dt = 0 = p(T)N - A_rNN_{mobile} \quad (16)$$

$$dN_2^*/dt = A_rNN_{mobile} \quad (17)$$

The TL intensity is proportional to the recombination rate, which itself is equal to the N atom detrapping rate in the steady state approximation:

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

$$I(t) \propto -dN_2^*/dt = \frac{1}{2}dN/dt = -p(T)N \quad (18)$$

Equation 18 can be integrated formally to include the temperature and time dependence of the detrapping rate constant $p(T(t))$.

Since

$$dN/dt = -2p(T(t))N \quad (19)$$

then

$$N(t) = N(t_0) \exp\{-\int_{t_0}^t 2p(T(t'))dt'\} \quad (20)$$

and

$$dN/dt = N(t_0) \exp\{-\int_{t_0}^t 2p(T(t'))dt'\} d/dt \left[-\int_{t_0}^t 2p(T(t'))dt' \right] \quad (21)$$

Equation 21 can be rearranged to give the final equation:

$$I(t) \propto 2N(t_0)p(T(t)) \exp\{-\int_{t_0}^t 2p(T(t'))dt'\} \quad (22)$$

If an activated process is assumed, $p(T(t))$ has an Arrhenius temperature dependence given by equation 23:

$$p(T(t)) = f(t, T) = \nu e^{-E_a/kBT(t)} \quad (23)$$

where ν is the frequency factor which corresponds to the number of collisions occurring; E_a is the activation energy barrier; and k_B is Boltzmann's constant. $T(t)$ is the time-dependent temperature ramp which was experimentally determined and recorded simultaneously with the TL curves. The intensity is then given by equation 24:

$$I(t) \propto N(t_0) \nu e^{-E_a/kBT(t)} \exp\{-\int_{t_0}^t 2\nu e^{-E_a/kBT(t')}dt'\} \quad (24)$$

This equation was used to calculate the theoretical glowcurves for given values of ν and E_a .

The histograms which were computed from the experimental TL intensity

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

data were matched with the best (least-squares) fits to the theoretical histograms calculated from equation 24. This equation was evaluated numerically using numerical integration and the experimentally determined temperature ramps to evaluate $T(t)$. The fits were performed for each peak by iteration of the ν and E_a to minimize the mean square errors.

Using three distinct diffusion activation enthalpies for the N/N_2 matrix, the three emission peaks of the histograms were each fit with a separate theoretical glowcurve. The three simulated curves, added together, normalized, and compared to the experimental data of the histogram in Figure 4b gave an excellent agreement. Table 2 lists the frequency factors and activation energies extracted from the intensity equation for each peak fit for the N/N_2 and N/D_2 matrices. The first glowcurve fit for the N/N_2 matrix yielded a ν value of 10 s^{-1} and a E_a of 24 K. The second peak yielded a ν of 10^3 s^{-1} and an activation temperature of 102 K. The final simulation gave a ν of 10^2 s^{-1} and 113 K for E_a . Figure 11 depicts the matches of the individual and cumulative computer simulated glowcurve fits with the experimental data showing an excellent agreement between the experimental and theoretical glowcurve shapes.

The same fit procedure was used for the N/D_2 matrix histogram from Figure 8b and is depicted in Figure 12. This histogram had only a single peak at 7 K and the extracted values for the frequency factor and activation energy were found to be 10^3 s^{-1} and 50 K, respectively.

The histograms in Figures 4 and 9 demonstrate the differences in the temperature ranges for the emissions in the different matrices. Figure 4b depicts the thermoluminescence glowcurve of the N/N_2 matrix where the emission is between 7 K and 23 K. Figure 9 shows the thermoluminescence glowcurves for the N/D_2 matrix samples. Here, the emission ranges from 3 K to 16 K. The reason for these temperature differences is most likely in the difference of the detrapping and recombination rates in the N/D_2 samples from the N/N_2 samples. The higher temperature emissions ($T > 8 \text{ K}$) may be due to solid nitrogen regions emitting inside the D_2 sample which formed as the D_2 left the matrix. This phenomenon is especially evident in the histogram of Figure 9c.

The three TL peaks at 8 K, 12 K, and 19 K in the N/N_2 matrix glowcurve from Figure 4b suggest three separate processes with their own activation enthalpies are occurring. This same glowcurve shape was observed previously by Hemstreet and Hamilton²⁰ and Brockelhurst and Pimentel²¹ indicating very

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

close agreement among all these results. Table 3 lists the temperatures of the peak emissions and the experimental conditions for the three individual studies. Hemstreet and Hamilton obtained two peaks, one at 16 K and another at 20 K. They missed the peak below 10 K because their initial sample temperature was above 10 K. Brockelhurst and Pimentel obtained three TL peaks at 10 K, 14 K, and 19 K. These investigators explained the three peaks in terms of three separate and distinct activation energies which are characterized in terms of the storage and diffusion of the detrapped nitrogen atoms. The differences in the temperatures obtained by the two studies was rationalized by Brockelhurst and Pimentel as being dependent on the deposition conditions.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Table 2. Comparison of results with those of Brockelhurst and Pimentel (BP) and Hemstreet and Hamilton (HH).

	Present Work	BP	HH
Amount of N ₂ deposited (mM)	.005	.01-.1	10-30
Deposition rate (mmoles/min)	.002	.01-.02	1-20
TL observed (K)	8	10	...
	12	15	16
	19	19	20

Table 3. N/N₂ detrapping parameters extracted from the fits to TL data from the derived intensity equation and the thermodynamic values calculated from them.

Matrix	Peak Temperature (K)	Frequency Factor (ν) (s ⁻¹)	Activation Energy (E _a) (K)
N/N ₂	8	10 ¹	24
	12	10 ³	102
	19	10 ²	113
N/D ₂	7	10 ³	50

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

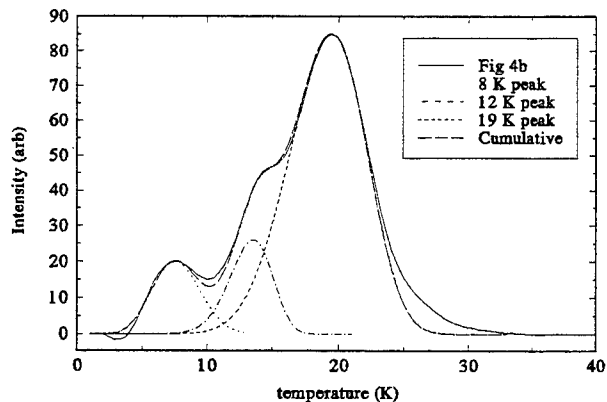


Figure 11. Comparison of the three first order theoretical glowcurves and their summation with the experimental data for the N/N_2 matrix. (---) is the histogram of Figure 4b while (...) is the fit to the 8 K peak, (-.-.) is the fit to the 12 K peak, and (- - -) is the fit to the 19 K peak. (-- --) is the summation of the three theoretical peaks.

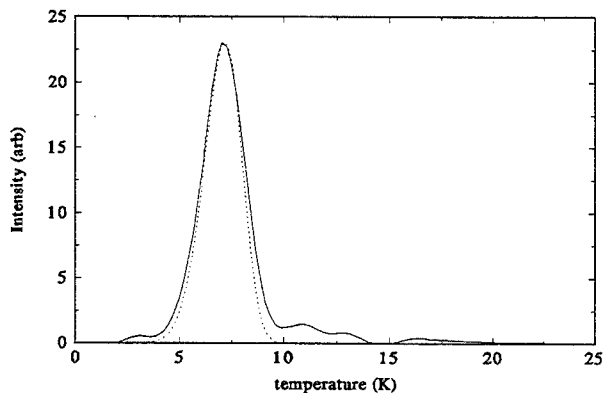


Figure 12. Comparison of the first order theoretical glowcurve with the experimental data for the N/D_2 matrix. (---) is the histogram of Figure 8b while (...) is the computer simulation.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Section C: Microscopic Model

Pure nitrogen freezes in a hexagonal-close packed (hcp) array, where each N_2 molecule is surrounded by 12 nearest neighbor molecules. The SMH also solidify in a close-packed array which could be either hcp or face-centered cubic (fcc)²² depending on how the sample is prepared. The low temperature vapor deposited matrices used in this study are probably composed of very small, defect-rich, micro-crystalline regions separated by even more disordered grain boundary regions.⁹ However, traditional models of solid phase diffusion are a useful place to begin the discussion. The observed TL in the matrix experiments is most likely dependent upon the diffusion of N atoms through the surrounding solid of N_2 or D_2 molecules.

There are three basic atomic mechanisms used to explain the diffusion in solids.²³ These are illustrated in Figure 13. The first mechanism involves an atom moving between the nearest neighbor molecules which surround it. This is known as the interstitial mechanism of diffusion (Figure 13a). The molecules near the atom are pushed aside by the atom, and there is a local distortion of the crystal structure. As the atom moves to its midpoint position, it encounters strong repulsive forces so that a considerable amount of energy must be supplied to the moving atom in order to enable it to squeeze past its neighbors. The source of this energy is the thermal vibrations of the molecules. The energy must be transferred to the atom to activate it sufficiently to overcome the repulsive forces of its surroundings. This configuration is the "activated state". Figure 13 a depicts the mechanism in three steps²⁴: (a) atoms in a crystal before the atomic jump; (b) atoms in an activated state; and (c) atoms in a crystal after the jump. The vibrating atom can collide with the neighboring molecules, which does not result in diffusion, or it can move into the spaces within the lattice. The frequency factors must be high enough to overcome the steric effects that prevent atomic movement between these molecules. Therefore the value of ν must be on an order of the lattice vibration, which is approximately 10^{10} to 10^{13} s⁻¹ with an activation energy on the order of 10^4 to 10^5 K. Because of these parameters, this mechanism is rarely known to happen.

The second mechanism of atom diffusion is known as the interchange mechanism and is the simplest mechanism by which an atom can move. It is simply the direct exchange of two (or more) nearest neighbor atoms (Figure

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

13b). This movement can happen in any direction since all the nearest neighbors are equivalent. The frequency factor and activation energy should be smaller than the interstitial mechanism and on the order of the self diffusion data.

The final mechanism is the vacancy mechanism whereby an atom originally in a normal lattice site moves into a neighboring vacant site (Figure 13c). This figure also shows the mechanism in three steps²⁴: (a) the vacancy in the initial position; (b) the activated state; and (c) the vacancy in the final position. This is the predominant diffusion mechanism for the lattice atoms in a close-packed array and the mechanism proposed to occur in the nitrogen matrices studied in this work. Vacancies are present in equilibrium concentrations at all temperatures above absolute zero. The diffusion values for the frequency factor and activation energy are generally also on the order of the self-diffusion values and can be compared to the experimental values used to fit the TL glowcurves.

The diffusion of atoms in matrices is, of course, temperature dependent. The Arrhenius equation can therefore also be written as follows:

$$k = \nu e^{-\Delta G^*/kT} \quad (25)$$

where ΔG^* is the free energy of activation taken from the thermodynamic definition of the free energy change, $\Delta G^* = \Delta H^* - T\Delta S^*$, where ΔH^* and ΔS^* are the changes of the enthalpy and entropy of activation, respectively. Substituting this relationship into the former equation yields equation 26:

$$k = \nu e^{\Delta S^*/k} e^{-\Delta H^*/kT} \quad (26)$$

Since the preexponential frequency factor is composed of a ν term and a $e^{\Delta S^*/k}$ term, the entropy of activation cannot be directly determined from the theoretical data. Also, the extracted frequency factors from the fits cannot necessarily be taken as the absolute values since there could be more than one diffusion mechanism taking place in the lattice. Furthermore, it is also difficult to determine exactly what the lattice looks like. Depending on the type of mechanism assumed and the kinetic order, the calculated ν factor varies greatly in magnitude.²⁵

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

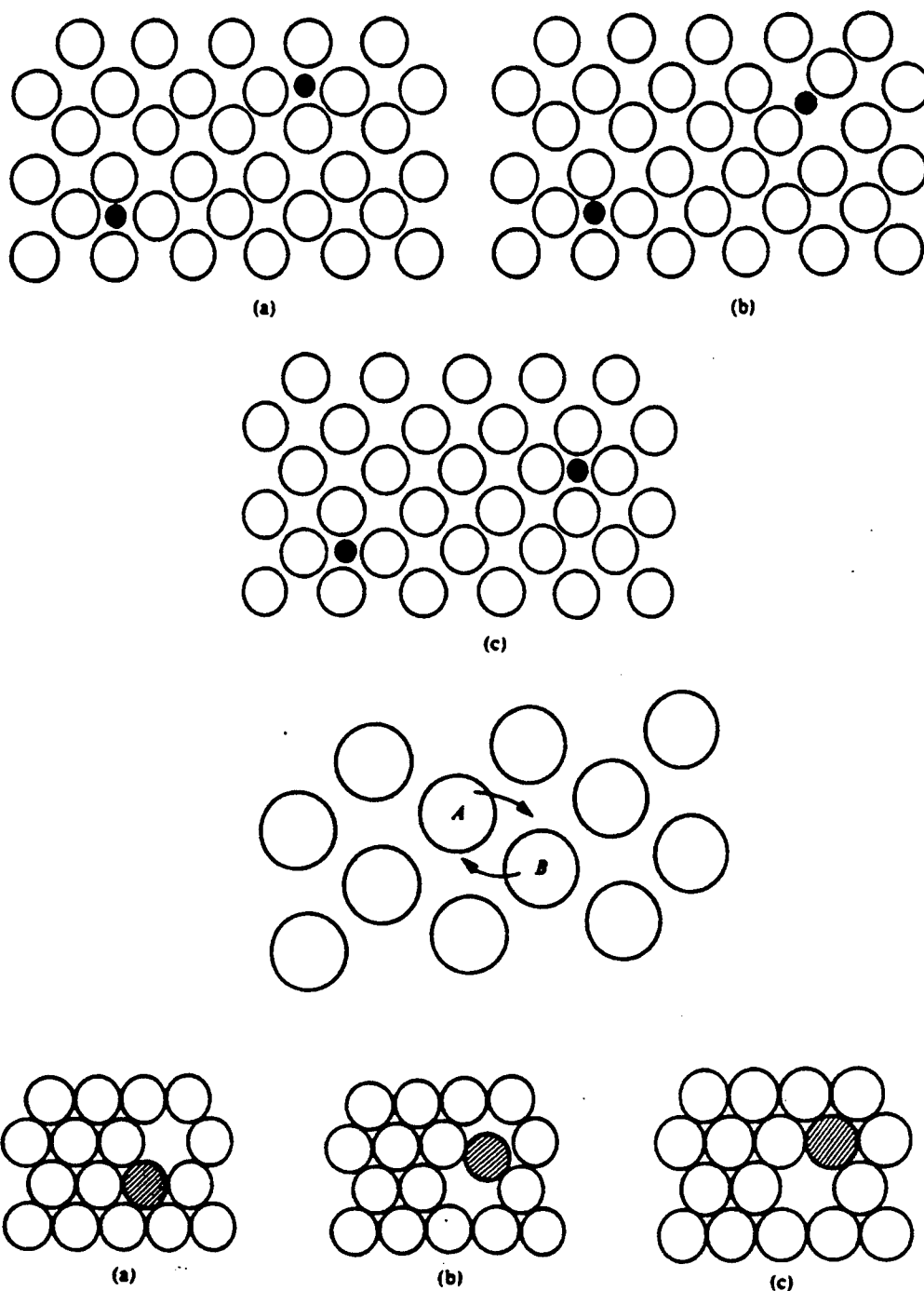


Figure 13. Three basic atomic diffusion mechanisms. The top shows the interstitial mechanism, the middle is the interchange mechanism, and the bottom depicts the vacancy mechanism.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

Conclusions

Atoms of N and O can be trapped in D₂ solids below 4 K where the N atoms must be present in order to see TL. The main species emitting light upon heating the matrices were N₂^{*}, N^{*}, O^{*}, and O₂^{*}. When the D₂ temperature rose above 4 K, recombination occurred rapidly over a short temperature range (4 K to 8 K) accompanied by atomic and molecular emission. No emissions were observed from reaction products of the matrix host with the guest radicals.

The N/D₂ matrix was found a favorable candidate for the scale-up experiments. The TL will serve as an internal thermometer for the doped D₂ solids since TL is a valid temperature probe.

The first-order kinetic model where the atom detrapping process is the rate determining step produced very good fits between the experimental and theoretical TL glowcurves. The values for the frequency factors and the activation energies were obtained by making several assumptions. This model assumes a homogeneous matrix with uniform temperature and concentration throughout the sample. But, recombination and emission released heat into the matrix causing the flashes of green light at 521.4 nm observed during the warm up. These thermal explosions may cause the local temperature in the matrix to rise for short periods of time. This could effect the TL intensity even though the temperature ramp is still rising smoothly since these local higher temperatures would cause more N atoms to detrapp from the matrix and become mobile. Also, the emission peaks were fitted with only a single detrapping energy and the system may have a distribution of energies or several detrapping energies because of the multiple peaks in the TL glowcurves. By refining some of these assumptions, a better fit may be found for the kinetic model.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

References

1. T.L. Thompson, ed., *Proceedings of the high energy density matter (HEDM) contractors' conference held 6-8 June 1993 in Woods Hole, MA*, PL-TR-93-3041, (USAF Phillips Laboratory, Edwards AFB, CA 93524-5000, 1993).
2. G.P. Sutton and D.M. Ross, *Rocket Propulsion Elements*, 4th edition, (Wiley, New York, 1976).
3. P.G. Carrick, PL-TR-93-3014, (USAF Phillips Laboratory, Edwards AFB, CA, 93524-5000, 1993)
4. M.E. Fajardo, M.E. DeRose, Y.K. Bae, to be published.
5. L. Vegard, *Nature* **113** (1924), 716.
6. J.C. McLennan and G.M. Shrum, *Proc. Roy. Soc.* **A106** (1924), 138.
7. A.M. Bass and H.P. Broida, eds., *Formation and Trapping of Free Radicals*, (Academic, New York, 1960).
8. E. Whittle, D.A. Dows, and G.C. Pimentel, *J. Chem. Phys.* **22** (1954), 1943L.
9. S. Cradock and A.J. Hinchcliffe, *Matrix Isolation*, (Cambridge University Press, Cambridge, 1975).
10. M.E. Fajardo, S. Tam, and M.E. Cordonnier, to be published.
11. M.E. Fajardo, S. Tam, T.L. Thompson, and M.E. Cordonnier, *J. Chem Phys.* **189** (1994), 351.
12. R.L. Brooks, *J. Chem. Phys.* **85** (1986), 1247.
13. M. Peyron, E.M. Hörl, H.W. Brown, and H.P. Broida, *J. Chem. Phys.* **30** (1959), 1304.
14. M. Peyron and H.P. Broida, *J. Chem. Phys.* (1959).
15. J.A. Meyer, D.W. Stetser, and D.H. Stedman, *J. Phys. Chem.* **74** (1970), 2238.
16. M. Peyron and H.P. Broida, *J. Chem. Phys.* **30** (1959), 139.
17. J.I. Steinfeld, *Molecules and Radiation: An Introduction to Modern Molecular Spectroscopy*, (MIT Press, Cambridge, 1985).
18. P.L. Kunsch and K. Dressler, *J. Chem. Phys.* **68** (1978), 2550.
19. M.E. Fajardo and V.A. Apkarian, *J. Chem. Phys.* **89** (1988), 4124.
20. R.A. Hemstreet and J.R. Hamilton, *J. Chem Phys.* **34** (1961), 948.

THERMOLUMINESCENCE STUDIES OF SIMPLE SPECIES IN
SOLID MOLECULAR HYDROGEN MATRICES

Janet Petroski

21. B. Brockelhurst and G.C. Pimentel, J. Chem Phys. **36** (1962), 2040.
22. P.C. Souers, *Hydrogen Properties for Fusion Energy*, (University of California Press, Berkeley, 1986).
23. P. Haasen, *Physical Metallurgy, 2nd edition*, (Cambridge University Press, Cambridge, 1986).
24. L.A. Girfalco, *Atomic Migration in Crystals*, (Blaisdell Publishing, New York, 1964).
25. V. Ausín and J.L. Alvarez Rivas, J. Phys. C: Solid State Phys. **5** (1972), 82.

Neural Network Approximation of Dynamic Viscosity-
Improving the Quality of Composite Structures

Richard M. Salasovich
Graduate Student
Department of Electrical and Computer Engineering

University of Cincinnati
Cincinnati, OH 45221

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

Phillips Laboratory

December 1995

**NEURAL NETWORK APPROXIMATION OF DYNAMIC VISCOSITY—
IMPROVING THE QUALITY OF COMPOSITE STRUCTURES**

Richard M. Salasovich
Graduate Student
Department Of Electrical And
Computer Engineering
University Of Cincinnati

Abstract

Various attempts are made to apply adaptive schemes to improve the quality of composite structures. On-line identification and approximation methods are explored to identify processing parameters and approximate dynamic viscosity in autoclave curing of composites. Problems encountered with each method are discussed. Dynamic viscosity of thermoset resin (generated by the Strother Dynamic Model) is approximated using a discretized multilayer neural network.

NEURAL NETWORK APPROXIMATION OF DYNAMIC VISCOSITY— IMPROVING THE QUALITY OF COMPOSITE STRUCTURES

Richard M. Salasovich
Graduate Student
Department Of Electrical And
Computer Engineering
University Of Cincinnati

Introduction

The application chosen for this adaptive control study was autoclave curing of composite materials. In the manufacture of polymeric composites, tows of fibers (which are usually composed of glass or graphite) are either laid side-by-side unidirectionally or woven into a fabric. Then, they are impregnated with a polymeric resin and partially cured in order to improve handling characteristics. Layers of this prepreg are stacked to form the shape of the desired part and enclosed in a vacuum bag. The assembly is then placed in an autoclave where heat and pressure are applied to lower the resin viscosity so that the polymer fully coats the fibers. This initiates an exothermic reaction in the resin which is called a cure. The process is complete when the polymeric resin is fully cured.

Current models for the curing of polymeric resin are very complicated and require precise definition of chemical reaction kinetics. Therefore, the goal of this study is to simulate proposed cure procedures for new resin systems using a simple model. Various approaches for on-line identification and approximation are explored: identification of a second-order linear model; identification of the Strother Model; approximation by the polynomial method; and approximation by a multilayer neural network.

Procedure

Approach 1: For the initial effort, an attempt is made to model the curing of a part in an autoclave as a second-order linear ordinary differential equation:

$$\dot{T}_{resin} = A_1 T_{resin} + B_1 T_{autoclave} \quad (1)$$

$$\dot{V} = A_2 V + B_2 T_{resin} \quad (2)$$

with T=temperature, V=viscosity, and A_1 , B_1 , A_2 , B_2 =time varying parameters. This approach is abandoned because it simplifies the problem to the point that it becomes meaningless. It was determined that equations (1) and (2) are unrealistic in modeling the cure of resin. Also, the A_2 parameter can not be readily sensed on-line. It must be inferred from other data (i.e. dielectric constant) and is somewhat imprecise, thereby introducing excessive error in an adaptive scheme.

Approach 2: The Strother Dynamic Model for thermoset resins is used in the following four approaches. For new resin materials, process development centers around resin viscosity and reaction rate for a given processing condition. Only after making a “good” test part does heat of reaction and optimum cure time become critical. The Strother Macrokinetics Model predicts resin reaction before the gel point. Isothermal gel time for a resin is given by:

$$GelTime = ae^{bT} \quad (3)$$

where T=temperature and a and b are matrix dependent constants. During any small time increment (dt) the Gel Fraction (amount of reaction that has occurred) can be expressed as:

$$GelFraction = G = \frac{dt}{ae^{bT}} \quad (4)$$

so:

$$\dot{G} = \frac{1}{ae^{bT}} \quad (5)$$

The dynamic case (temperature changing over time) is treated as a series of isothermal steps. The gel fraction for each step is calculated. The reaction fraction (total amount of reaction that has occurred) is the sum of these gel fractions over the dynamic range.

$$ReactionFraction = \sum GelFraction \quad (6)$$

When the reaction fraction equals 1 then the matrix has gelled. The viscosity profile of a resin system heated linearly at 2-3°C/min is generally a U-shaped curve, as shown in Figure 1 of Appendix A. For the left leg of the U, little chemical reaction has occurred and the classical inverse relationship of temperature and viscosity for the starting resin is observed. The right leg of the U represents a more complex case. Significant reaction has occurred, therefore the viscosity increases. At the same time system viscosity is decreasing due to increasing temperature. Strother proposes the total resin viscosity be modeled as a simple mixture:

$$Ns = K1e^{\frac{K2(1-RF)}{T}} \quad (7)$$

$$Nr = K3e^{K4RF/T} \quad (8)$$

$$N_{total} = Ns + Nr \quad (9)$$

where Ns =starting resin viscosity; Nr =reacted product viscosity; N_{total} =total resin viscosity; RF =reaction fraction; T =temperature; and $K1, K2, K3, K4$ are resin dependent constants. These equations predict the system viscosity at any temperature and reaction fraction up to the system gel point. Applying them beyond the gel point produces erratic results. The Strother Model is coded in MATLAB as shown in Appendix A. This generates data needed for the adaptive schemes.

Using the Strother Model, an attempt is made to use on-line identification techniques to estimate the parameters $a, b, K1, K2, K3, K4$. However, the system can not be linearly parameterized as:

$$\dot{y}(t) = -ay(t) + br(t) \quad (10)$$

because the states are functions of logs and exponents.

Approach 3: Next, on-line approximation is attempted using the polynomial method. As seen from Figures 2 and 3 of Appendix A, the inputs to the polynomial method (temperature and reaction fraction) continually increase. In order to guarantee persistency of excitation so that the parameters will converge to their real values, the cure cycle of the autoclave has to be driven unrealistically hard. Even if this were implemented, parameter convergence might not occur because the cure process is such a slowly migrating one. Therefore, the polynomial method is abandoned.

Approach 4: On-line approximation is attempted using a multilayer neural network. However, this is not realizable because reaction fraction can not be calculated continuously with the adjustable weights.

Approach 5: Finally, an off-line discretized multilayer neural network is implemented to approximate resin viscosity as a function of temperature. The equations for a 2 layer feedforward neural network are given in Appendix B, along with the neural network architecture of Figure 4. These are coded in MATLAB as shown in Appendix B. Inputs (temperature and reaction fraction) are generated using the Strother Model shown in Appendix A. Data sets of 12 and 113 exemplars were used to train networks with 3, 6, 9, 12, and 15 neurons. The output is viscosity. The magnitudes of the inputs and outputs vary by orders of magnitude (temperature=50-175°C; reaction fraction=0-1; viscosity=3-15,000 poise). These values have to be normalized to between zero and one so that the network can be trained

effectively. Otherwise, training takes an excessively long time. Even with data normalization, there error never converged to < 0.05 with up to 14 hours of training.

Results

An example of Viscosity vs. Temperature and Error vs. Epoch is given in Figure 5 of Appendix B. Networks with 3, 6, 9, 12, and 15 neurons are trained to ten-thousand epochs with 12 and 113 data sets. Error values are given in Table 1. Error is seen to decrease with increasing number of neurons, however it also increases the training time.

# Data Sets	# Neurons	Error
12	3	1.92
12	6	1.72
12	9	1.62
12	12	1.49
12	15	1.42
113	3	22.68
113	6	16.58
113	9	15.98
113	12	14.92
113	15	14.85

Table 1. Error Values After 10,000 Epochs

Conclusion

Various approaches for on-line identification and approximation for autoclave curing of composites are explored: identification of a second-order linear model; identification of the Strother Model; approximation by the polynomial method; and approximation by a multilayer neural network. The Strother Model is coded to generate data sets to train a discretized 2-layer feedforward neural network. The network is trained with 3, 6, 9, 12, or 15 neurons and data sets consisting of 12 or 113 exemplars. It is determined that increasing the number of neurons improves the error. The trained network can be used to predict dynamic viscosity during the autoclave cure cycle.

References

- LeClair, S. R. "Qualitative Process Automation." Int. J. Computer Integrated Manufacturing, 1989.
- Strother, R. K. "Elementary Empirical Process Modeling For Thermoset Resins." Proceedings of the 31st International SAMPE Symposium and Exhibition, 1986.
- Zurada, J. M. Introduction To Artificial Neural Systems. West Publishing Co., 1992.

Appendix A

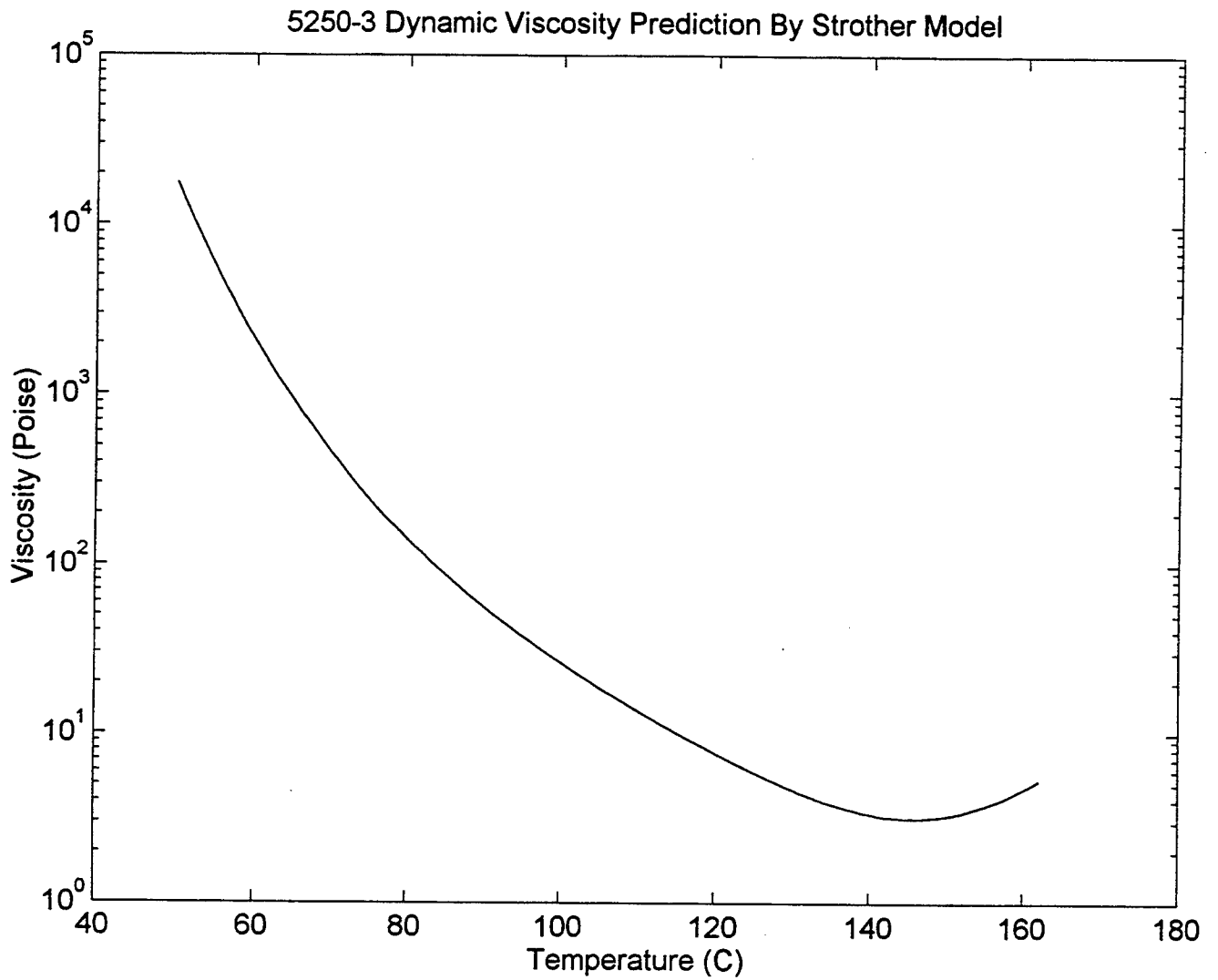


Figure 1

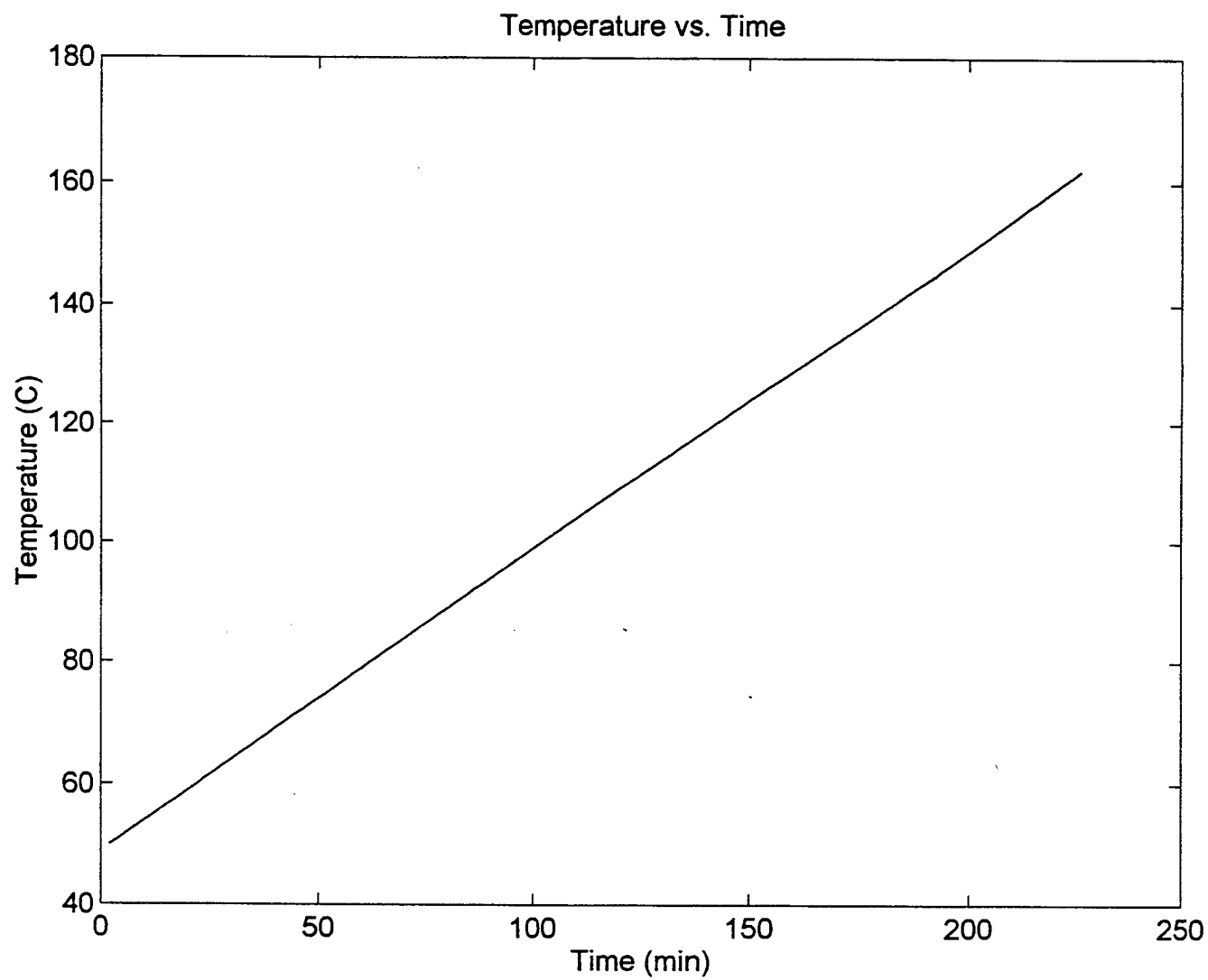


Figure 2

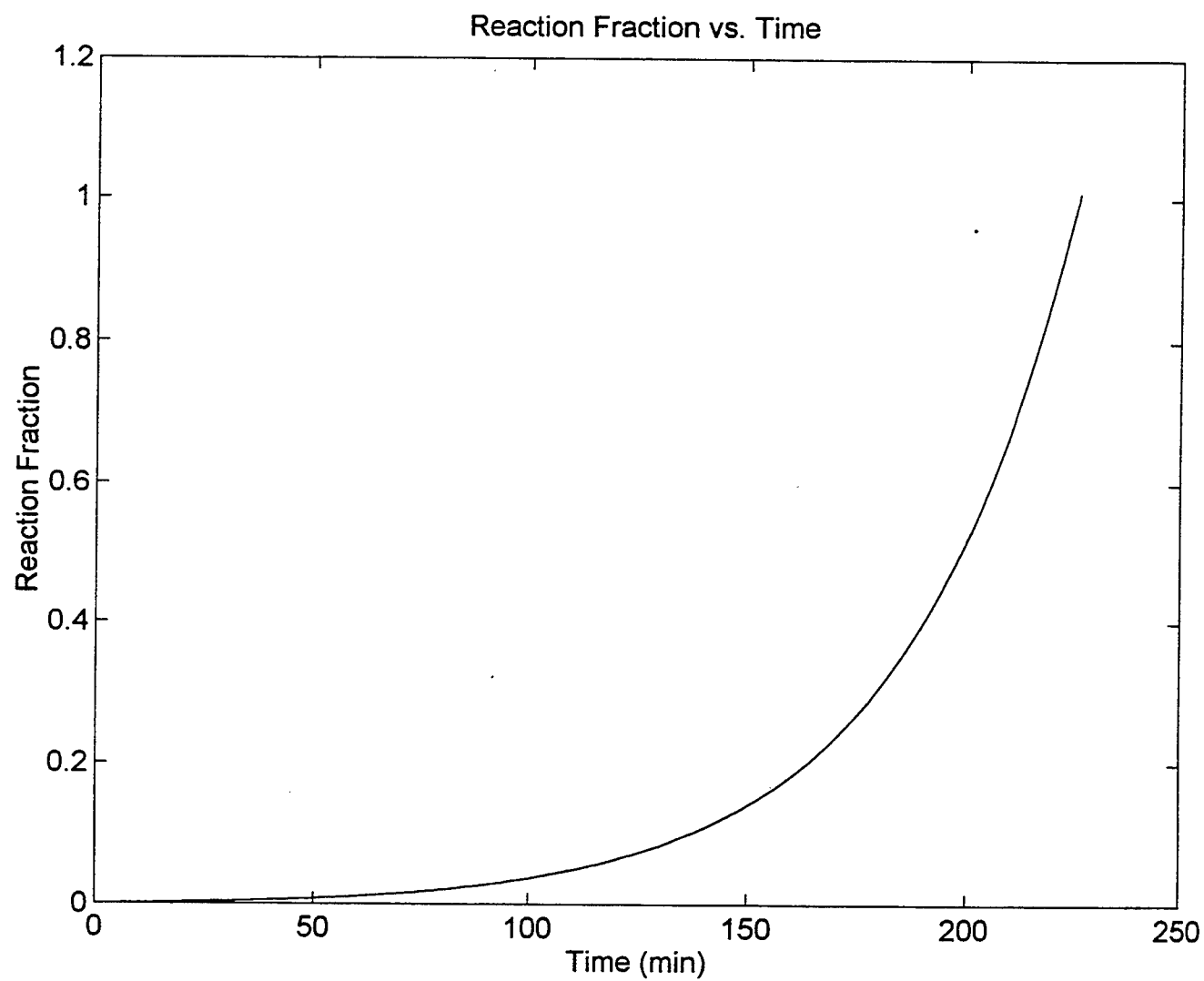


Figure 3

```
% Viscosity Prediction By Strother Model
```

```
clear;
a=83333; b=-0.0512; R=2;
K1=0.0579; K2=631.4423; K3=1.2649; K4=229.0746;
RF=0;
TT1=[50,62.5,75,87.5,100,112.5,125,137.5,150,162.5,175];
VV1=[11000,1700,200,45,17,6,3,2,2.5,7,60];
p=polyfit(TT1,VV1,10);
T=50:175;
VV2=polyval(p,T);

for count=1:size(T,2)
    if RF<1
        RF=RF + 1/(a*exp(b*(T(1,count) + R/2)));
        junk(1,count)=count;
        junk(2,count)=T(1,count);
        junk(3,count)=RF;
        junk(4,count)=RF/T(1,count);
        junk(5,count)=VV2(1,count);
        Nm=K1*exp((K2*(1-RF))/T(1,count));
        Nr=K3*exp((K4*RF)/T(1,count));
        Ns(1,count)=Nm + Nr;
    end
end
semilogy(junk(2,:),Ns(1,:))
title('5250-3 Dynamic Viscosity Prediction By Strother Model')
xlabel('Temperature (C)')
ylabel('Viscosity (Poise)')
junk1=[junk(1,:)',junk(3,:)',junk(4,:)'];
junk2=[junk(1,:)',junk(5,:)'];
junk3=[junk(2,:)',junk(3,:)',Ns(1,:)'];
```

```
function xdot=viscous(T,x)
```

```
a=83333; b=-0.0512;
```

```
xdot=[1/(a*exp(b*T))];
```

Appendix B

2-Layer Feedforward Neural Network Equations (3 Neurons)

Input Data Set : $\begin{bmatrix} T \\ RF \end{bmatrix}$ (2 x 113) Output Data Set: $[V]$ (1 x 113)

$$ybar = A \begin{bmatrix} T \\ RF \\ 1 \end{bmatrix} \quad (1 \times 3)$$

$$y = \begin{bmatrix} \sigma(ybar) \\ 1 \end{bmatrix} \quad (1 \times 4)$$

$$\sigma(ybar) = \frac{2}{1 + e^{-\lambda ybar}} - 1 \quad (1 \times 3)$$

$$zbar = Cy \quad (1 \times 4)$$

$$V = \sigma(zbar) \quad (1 \times 1)$$

Adjust Weights:

$$\Delta_{output} = \frac{1}{2}(V_{actual} - V)(1 - V^2) \quad (1 \times 1)$$

$$C = C + \delta \Delta_{output} y \quad (1 \times 4)$$

$$\Delta_{hidden} = \frac{1}{2}(1 - y^2)(C \Delta_{output}) \quad (1 \times 4)$$

$$A = A + \delta [\Delta_{hidden} [T \quad RF \quad 1]]^T \quad (3 \times 3)$$

Use these weights to step through all input data sets

$$Error = \sum |V_{actual} - V| \leq 0.05 \quad (1 \times 1)$$

Neural Network Architecture

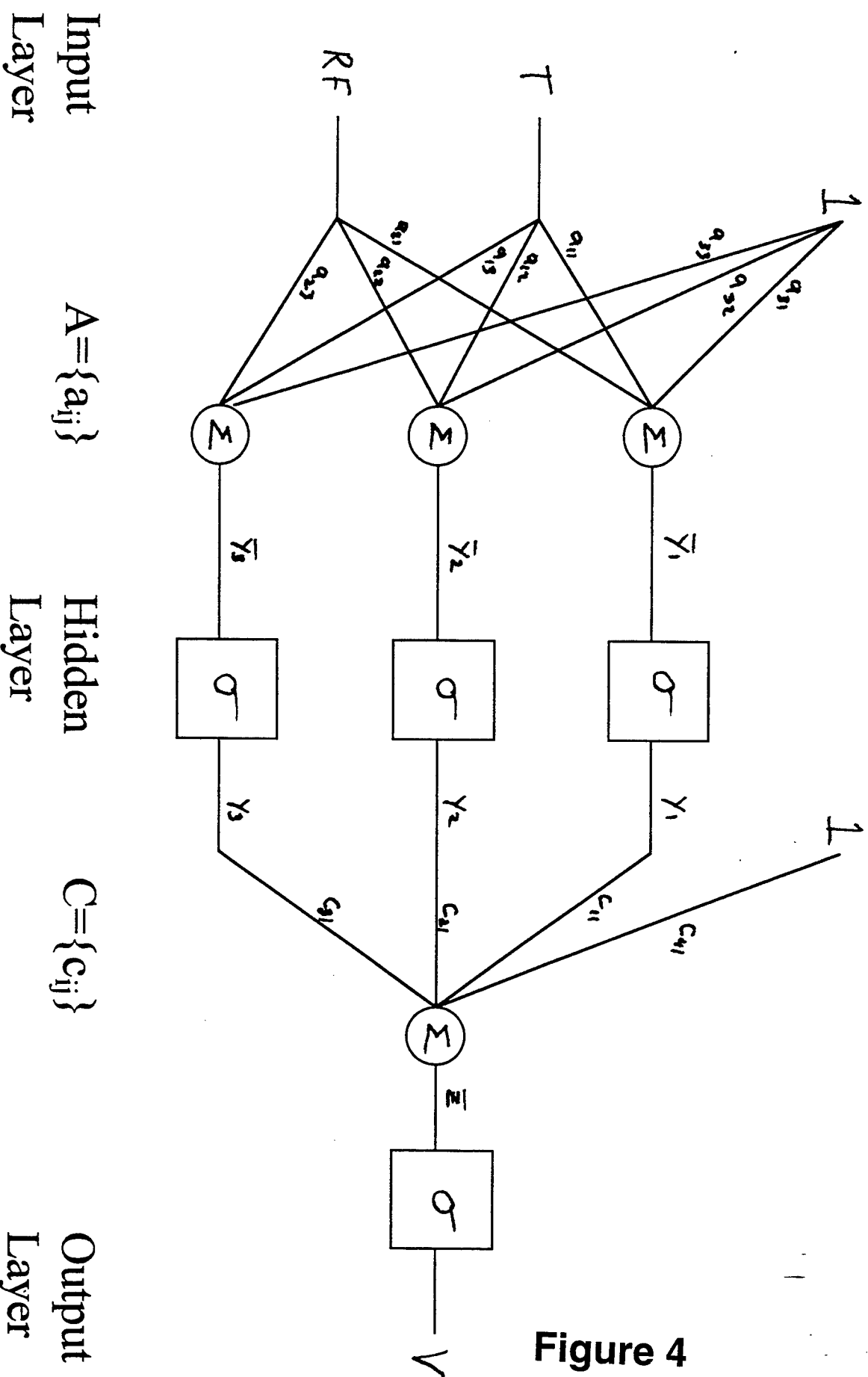


Figure 4

% Multilayer Feedforward Neural Network

```

clear;
load data2.ascii;      % exemplars
z = data2;
P = 12; % number of exemplars
addz = ones(1,P);
z = [z;addz]; % inputs with bias input
load d2.ascii; % desired output
I = 2; % number of inputs
J = 3; % number of hidden
K = 1; % number of outputs
n = 0.01; % learning rate
errmax = 0.05;
lam = 1.0;
V = (rand(I+1,J)-0.5)*2; % weights from input to hidden
W = (rand(J+1,K)-0.5)*2; % weights from hidden to output
error = 100.0 % initialize error
k1 = 1; % indicates exemplar to use for training
epoch = 0;
while error > errmax;
    for cou = 1:1000
        y = V' * z(:,k1); % hidden neuron output
        y = ( 2.0 ./ ( 1 + exp(-lam .* y)) ) - 1;
        y(J+1,1) = 1; % bias for hidden neuron
        o = W' * y; % network output
        o = ( 2.0 ./ ( 1 + exp(-lam .* o)) ) - 1;
        delout = 0.5 .* (d2(:,k1) - o) .* (1 - o.^2);
        delhid = (0.5 .* (1 - y.^2)) .* (W * delout);
        W = W + (n .* (delout * y'))';
        V = V + (n .* (delhid(1:J) * z(:,k1)'))';
        if k1 >= P
            k1 = 0;
        end
        epoch = epoch + 1;
        k1 = k1 + 1;
        error = 0;
        k2 = 1;
        while k2 <= P;
            y = V' * z(:,k2); % hidden neuron output
            y = ( 2.0 ./ ( 1 + exp(-lam.* y)) ) - 1;
            y(J+1,1) = 1; % bias for hidden neuron
            o = W' * y; % network output
            o = ( 2.0 ./ ( 1 + exp(-lam .* o)) ) - 1;
            out(k2) = o;
            error = error + sum(abs( o-d2(:,k2) ));
            k2 = k2 + 1;
        end
        track(epoch) = error;
    end
    plot(track)
    sum(abs(d2-out))
    out
    epoch
end
out

```

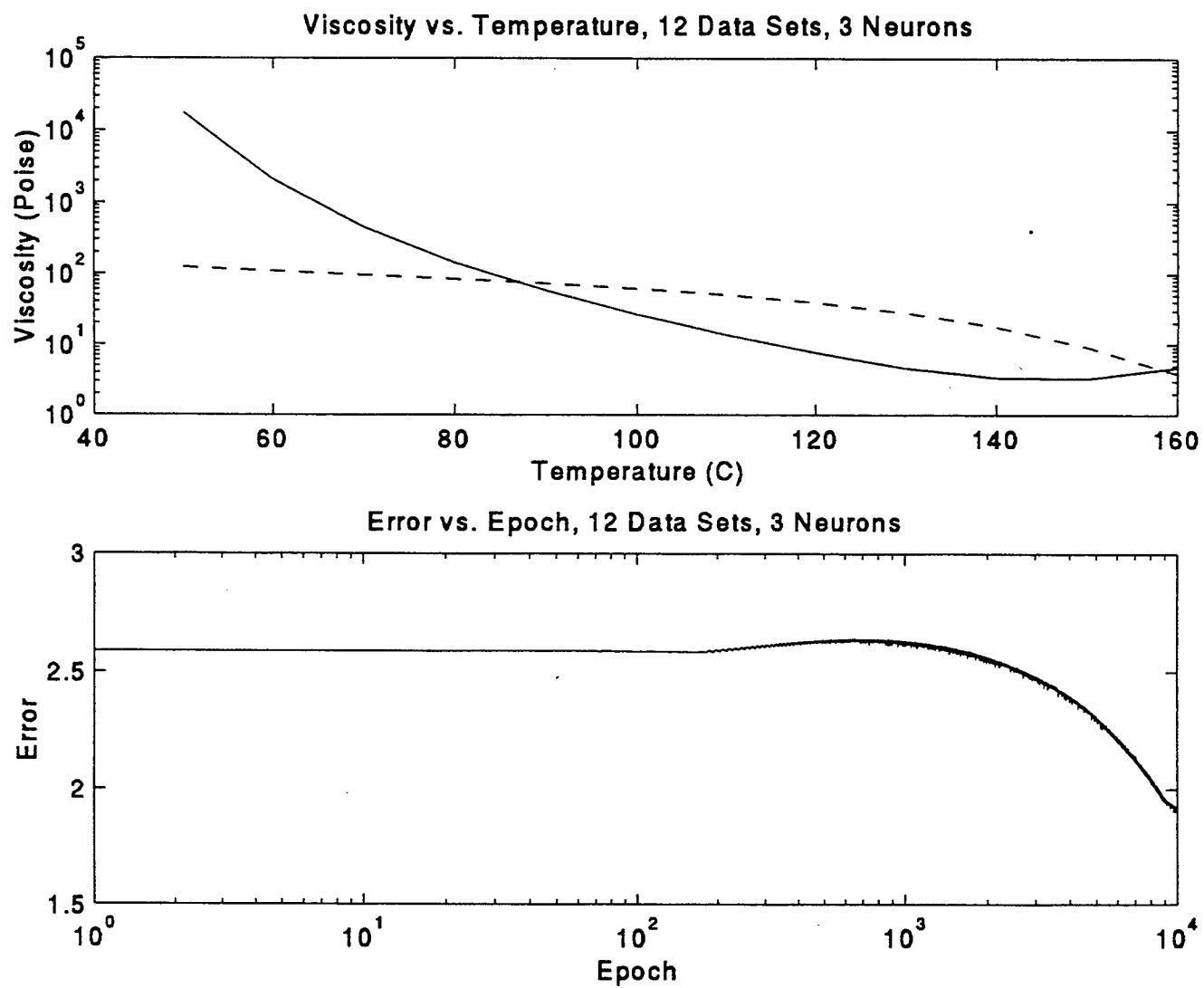


Figure 5